Replication Studies

Learning vocabulary through meaning-focused input: Replication of Elley (1989) and Liu & Nation (1985)

Stuart Webb Faculty of Education, Western University, Canada
swebb27@uwo.ca

There has been a great deal of research on first language (L1) and second language (L2) learning through meaning-focused input since Nagy, Herman & Anderson’s (1985) seminal study of incidental vocabulary learning through reading. Two strands of research within this area are incidental vocabulary learning through listening and guessing from context. This article discusses widely-cited studies from each of these areas – Elley (1989) and Liu & Nation (1985) – that deserve to be replicated. Both studies made important contributions to the field, were original in their designs, and advanced our understanding of how vocabulary is learned in context. The benefits of replicating these two studies are described and several suggestions are made for how the replications might be completed.

1. Introduction

Research has indicated that L1 words can be learned incidentally through reading (e.g. Jenkins, Stein & Wysocki 1984; Nagy, Herman & Anderson 1985) and listening (Elley 1989; Robbins & Ehri 1994). Motivation for this research has been the hypothesis that the vast majority of L1 words are learned incidentally. Studies of incidental vocabulary learning through reading provide support for this as they have consistently shown that words are learned gradually through repeated encounters in context. Research has also shown that L2 vocabulary is also learned incidentally through reading (e.g. Pitts, White & Krashen 1989; Day, Omura & Hiramatsu 1991; Horst, Cobb & Meara 1998) and listening (Brown, Waring & Donkaewbua 2008), and that gains are a function of frequency (e.g. Horst, Cobb & Meara 1998; Webb 2007). However, a lack of L2 input may limit the extent of L2 incidental vocabulary learning to the high frequency words (Cobb 2007), leading some researchers to argue that a large proportion of L2 words may be learned through instruction rather than incidentally (Laufer 2001, 2003; Webb 2008a).

Although there is an abundance of research that has investigated incidental vocabulary learning through reading, there have been few studies that have looked at the potential for learning L1 and L2 words through listening. The best known study of incidental vocabulary learning through listening was conducted by Elley (1989). Elley’s research was innovative in that it was the first study to show that vocabulary could be learned incidentally through
listening, and that teacher explanations of word meanings during storytelling may increase the size of these gains. One reason why research on vocabulary learning through listening is lacking may be that it is more difficult to learn words through listening than through reading (Brown, Waring & Donkaewbua 2008). Thus, perhaps the lack of research in this area may be due to an absence of statistically significant findings indicating that vocabulary is consistently learned through listening.

Listening plays an important role in vocabulary growth; between 5,000 and 6,000 word families are learned through listening by pre-reading children (Goulden, Nation & Read 1990; Biemiller 2005). Because of the significance of Elley’s 1989 study, and the paucity of research in this area, the best place to start with new studies of vocabulary learning through listening is with replications of the original study. Replications would throw some light on the validity and reliability of the previous results, and may provide some indication of the extent to which the results may be generalized to other contexts.

Because vocabulary learning gains made from encounters in meaning-focused input are relatively small, another line of research has focused on how we might increase the potential for vocabulary growth through reading and listening. Studies of guessing from context have indicated that by focusing greater attention on unknown words encountered in context, the chance of learning words increases. Guessing from context is the most commonly used strategy for learning words from meaning-focused input and is a frequently researched area of vocabulary learning (e.g. Hulstijn 1992; de Bot, Paribakht & Wesche 1997; Fraser 1999; Paribakht & Wesche 1999). Liu & Nation’s (1985) study is unique in that it looked at how lexical coverage affects the ability to successfully infer word meanings from written text. Lexical coverage refers to the percentage of words that are known in a text. Research investigating lexical coverage has shown that as the proportion of known words in a text increases beyond 90%, comprehension is also likely to increase (Laufer 1989; Bonk 2000; Hu & Nation 2000; Schmitt, Jiang & Grabe 2011; van Zeeland & Schmitt 2013a). These studies have in turn contributed to greater focus in the research literature on the number of words necessary to reach the lexical coverage points associated with comprehension of spoken discourse (e.g. Nation 2006; Webb & Rodgers 2009a, 2009b; Rodgers & Webb 2011) and written (Hirsh & Nation 1992; Adolphs & Schmitt 2003; Nation 2006; Webb & Macalister 2013), as well as the potential to incidentally learn vocabulary through comprehensible input (Webb 2010).

Liu & Nation’s (1985) study indicated that lexical coverage affects the potential to successfully guess words from context; there is a better chance of successfully inferring unknown words in texts with a higher lexical coverage than passages with a lower lexical coverage. This finding has great value as it supports an extensive L2 learning approach that involves reading and listening to large quantities of comprehensible input as a means to increase vocabulary size. The importance of this finding and the lack of research in this area provide justification for approximate and conceptual replications of the original study. Replication of the research would help to verify the original findings and may provide greater clarity as to how lexical coverage affects guessing from context.

This article presents the case for replication of two important studies of vocabulary learning through meaning-focused input. Justification for replication of the original studies is outlined in the following sections. For each study, the methodology is first described and
then approaches to replication are suggested. Finally, the conclusion sums up the reasons why the research remains important to the field and why replication of these papers is warranted.


2.1 Background to the study

Liu & Nation (1985) is one of many studies that have looked at L2 learners’ ability to guess words from context. For example, one strand of research has focused on the procedure for guessing from context (Clarke & Nation 1980; Nation & Coady 1988; Nation 1990, 2001, 2013). Another has looked at the types of contextual clues that can be used to help derive word meaning from context (Haastri 1985, 1987, 1991; de Bot et al. 1997; Nassaji 2003). A third strand of research has indicated that instruction and proficiency affect guessing from context; instruction has a positive effect on successful inferencing, but the effects of instruction may vary according to proficiency level (Fukkink & de Glopper 1998; Kuhn & Stahl 1998; Walters 2004). A fourth strand has looked at the process of guessing from context through introspective and retrospective L2 learner data (Haastri 1985; de Bot et al. 1997; Nassaji 2003). Finally, a fifth strand on guessing from context has examined how inferring word meaning is affected by the information that is present in the context (e.g. Haastrup 1985; Sternberg 1987; Hulstijn 1992; Dubin & Olshtain 1993; Haynes 1993; Webb 2008b). Liu & Nation’s (1985) research is most closely related to this latter line of research.

Liu & Nation (1985) is the only study to examine the effects of lexical coverage on inferring the meanings of unknown words from context during reading. Native and non-native teachers of L2 English took part in the study. The participants read versions of two passages and were asked to write the meanings of pseudowords that had replaced lower frequency words in the texts. The meanings could be expressed in either the L1 or L2. The passages varied in length and lexical coverage. Two versions of one passage were provided in the appendix. However, information about the other passage is not sufficient for use in a replication. In one version of the passages, the lexical coverage was about 90%; approximately one word in ten was unknown. In these texts, all of the words that were less frequent than the most frequent 2,000 word families in West’s (1953) General Service List were replaced with pseudowords. The pseudowords were of a similar length to the words they replaced (prined – traced, radalosumic – philosophical), had suitable derived forms (benalment – dominance, alamution – environment), and included the same inflections (antelasing – prevailing, alliranes – attitudes). The second versions of the passages were expanded versions of the first texts. The procedure for creating these texts involved replacing all of the words outside of the General Service List and the American University Word List (Praninskas 1972) with pseudowords until there were the same numbers of pseudowords as in the first text. This resulted in texts that covered the same content as the first versions, but had a much higher lexical coverage. Lexical coverage in the expanded versions was approximately 96%; one in
every 25 words was unknown. The participants always completed the shorter version of the two texts first.

The participants’ responses were scored by three raters on the degree to which they accurately represented the meaning of the replaced words in those contexts. The raters scored responses for accuracy as 100%, 80% and 0% correct, and examples of responses that achieved these scores were provided. Responses that were given scores of 100% and 80% were counted as correct answers. No information was provided on how the three raters reached agreement on the overall score for each response.

Analysis of the data provided partial support for the hypothesis that words are more likely to be successfully guessed in context when there is a higher proportion of known words in a text. The descriptive statistics suggested that the participants were able to successfully derive the meanings of more pseudowords for one version that had higher lexical coverage, but not for the other. The results also revealed that the scores were higher in the version that had higher lexical coverage for nine out of ten pseudowords that were the same in both versions of the texts. However, inferential statistical analysis was not completed and so it is not clear to what extent the findings were due to chance.

2.2 Approach to replication

This study deserves replication because it employed an innovative approach to investigating guessing from context, the intervention indicated that lexical coverage is a factor in successful guessing, it is a frequently cited study (248 citations listed by Google Scholar on 28 May 2014), and because of the attention that lexical coverage has recently received in a number of studies looking at comprehension of spoken (Bonk 2000; van Zeeland & Schmitt 2013a) and written discourse (Laufer 1989; Hu & Nation 2000; Schmitt, Jiang & Grabe 2011). It may be most useful to conduct a series of replications that investigate the same research questions with a methodology that is modified slightly to shed greater light on the original study’s outcomes.

A series of replications would be useful to determine the reliability of the earlier findings and the degree to which the results might be generalized to other contexts. Replication might involve the manipulation of key variables such as text type and proficiency level that have pedagogical significance and may affect guessing from context. One methodological variation that would be useful to adopt in any replication is the approach taken to manipulating coverage that has been used in studies investigating how lexical coverage affects comprehension (Hu & Nation 2000; Schmitt, Jiang & Grabe 2011; van Zeeland & Schmitt 2013a). In these studies, texts with varying lexical coverage figures all had the same number of running words. In Liu & Nation, the two texts were of different lengths. Because it could be argued that the additional text in the longer versions may include information that could help to infer the meanings of target words, it would be better to always use the longer version of the text but to vary the coverage between versions. This would also allow the possibility of creating versions with more than the two levels of coverage in the original study. Examining the degree to which target words were successfully inferred in versions that had 90%, 95%, 98% and 100% lexical coverage would provide more precise
findings, and help to substantiate the earlier research. A second methodological innovation that might be included in any replications is using an additional delayed post-test. This would shed light on the degree to which the original findings are valid at different retention intervals.

In one approach to approximate replication, it would be useful to examine the extent to which words are correctly inferred from context in a range of different text types. Because there are a variety of text types (graded readers, short authentic reading passages, spoken texts such as dialogues and monologues) used for language learning, investigating how guessing from context is affected by lexical coverage in different texts would help us to better interpret the earlier findings and indicate the extent to which they might be generalized to other texts. Because learners are exposed to a wide variety of texts, it is particularly important to examine the effects of lexical coverage on guessing from context in different texts to strengthen the reliability of the original findings. In a second approximate replication, it would be useful to look at how guessing from context is affected by lexical coverage for groups of learners at different levels of proficiency. Liu & Nation’s study included participants from a range of levels but did not examine the extent to which such learners were able to successfully infer words in the study. Research has shown that L2 learners with a larger vocabulary size tend to be more successful at learning words incidentally through reading (Zahar, Cobb & Spada 2001; Tekmen & Daloglu 2006). Therefore, a replication of the Liu & Nation study that manipulates the language learning profile of the participants would help to determine the degree to which the earlier findings might be generalized across different groups of learners.

Examining the extent to which beginner, intermediate and advanced learners were able to infer the meanings of words in texts at different levels of lexical coverage might indicate the degree to which the original findings are valid for different learning profiles. Moreover, if learners at different levels vary in the degree to which they can successfully infer the meanings of words at different lexical coverage levels, it might better explain the results of the earlier studies of incidental vocabulary learning through reading.

There would also be value in undertaking a conceptual replication investigating the effects of coverage on guessing from context. This might involve having individual participants do a think-aloud protocol while completing the same guessing from context tasks outlined above. The research would examine the same questions as in the original study and include the same design modifications suggested above. However, a qualitative element built into a replication might further shed light on whether lexical coverage hides any useful data such as which words in the texts are outside the coverage of known words, what is the contribution to the text that these unknown words make, and how accessible are these unknown words to the learner?


3.1 Background to the study

This study explored the effects of repeated listening of a story on L1 incidental vocabulary learning, as well as the effects of teacher explanation of target words when the story was read aloud in two experiments. The research expanded on some of Elley’s earlier results,
found during piloting in his well-known book flood studies (Elley & Mangubhai 1983; Elley 1991). Elley hypothesized that children incidentally learn many words through listening to stories and that teacher explanation of difficult vocabulary in a story increases the amount of learning.

The first experiment investigated the extent to which young children incidentally learned vocabulary through listening to a story read three times over one week. Seven experienced teachers and 157 seven-year-old children from seven classes in seven schools in New Zealand took part in the study. The students formed the experimental group. There were no control or comparison conditions in this experiment. The story, *Gumdrop at Sea* (Biro 1983), was selected by the teachers because it was judged to be likeable, included many attractive pictures and contained at least 20 words that were thought to be unknown to the children. The story was read at a normal pace over approximately nine minutes. In a pilot study, four of the target words (direct, impatient, squeezed, disaster) were found to be known by many children with a similar learning profile as the participants. These words were replaced in the story with lower frequency synonyms (regulate, irritated, crammed, calamity). Twenty target items were selected for the study and these items and their corresponding learning gains are presented in the study.

Elley hypothesized that six textual factors may affect the vocabulary learning gains. These were: the number of occurrences of the target items in the text, the number of times the target items were included in pictures within the text, the information present in the story that could be used to infer the target word meanings, the importance of the target words to the development of the storyline, imageability of target items, and the students’ familiarity of the target meanings. Three teachers who did not participate in the study rated the last four factors on a six-point scale. The data was also analysed according to the participants’ pre-test scores to determine whether prior vocabulary knowledge affected learning gains.

The study used a quasi-experimental pre-test, post-test design. Vocabulary knowledge was measured using a 20-item multiple-choice test with the target word provided in both aural and written modes. For half of the items, the choices were made up of four pictures (the key and three distractors). For the other ten items, the choices were definitions (the keys were synonyms and three distractors). Several examples of the test items were provided in the appendix.

The analysis revealed that knowledge of all but six of the target items increased by more than 10%. There was a mean increase of 15.4% with gains ranging from –7.6% to 41.4%. Inferential analysis of the difference between pre-test and post-test scores was not provided. Multiple regression analysis indicated that three of the six factors (the number of occurrences of the target items in the text, the number of times the target items were included in pictures within the text, the information present in the story that could be used to infer the target word meanings) had a significant positive medium or high correlation with learning, and that four of the factors combined accounted for 53% of the variance.

Elley reported that the lack of a control group, the absence of delayed post-tests, and the use of only one story were limitations to the first experiment. The second experiment addressed these limitations by examining learning through listening to two other stories with post-tests at two retention intervals (one week and three months after the treatment) and included a control group to ensure that learning did not occur from taking the test of vocabulary
knowledge. It also further investigated the relationship between the textual variables and learning, and looked at the effects of teacher explanation of target words on vocabulary learning.

A between-participants design was used in the second experiment and 178 eight-year-old students from eight intact classes were assigned to two experimental and one control group. Eight experienced teachers participated; a different teacher taught each of the eight classes. Group A heard the first story read with explanations of the meanings of the target words. The meanings of the target words could be explained in three ways: by providing definitions, using role play, and by pointing to a picture. Group B heard the same story without the explanations using the same procedure as in the first experiment. The groups were counterbalanced for the second story. The test was administered to a third group at the same retention intervals, but it did not listen to the stories. Its role was to control for a possible learning effect from taking the test.

The descriptive statistics indicated that listening to the stories without explanation contributed to incidental vocabulary learning although the size of the scores indicated that the amount of learning may vary from text to text. This result supported the findings of the first experiment. The results also suggested that teacher explanation may contribute to much larger vocabulary learning gains than listening without teacher support for vocabulary learning. The descriptive data also indicated that the participants with the lowest scores on the pre-test made the greatest gains. However, because participants with higher pretest scores could learn fewer words than those with lower pre-test scores, it is necessary to examine relative learning gains rather than actual learning gains in this analysis (Shefelbine 1990; Horst, Cobb, & Meara 1998). However, the absence of inferential analysis and effect sizes in all of the comparisons between the groups makes it difficult to clearly determine the validity of these conclusions. Multiple-regression analysis indicated that all six variables had medium to large correlations with vocabulary learning in one text. Results of the analysis with the other text were not provided.

3.2 Approach to replication

There are several reasons why replication of this study is warranted. First, the study has had a large impact on the field and is widely cited as evidence of incidental vocabulary learning occurring through listening (739 citations listed by Google Scholar on 1 June 2014). Second, there are many positive features of the research design. Elley provides sufficient information about the methodology (the texts, target items, test items and teachers’ explanations) to allow replication for two of the three stories used in the study. His use of a pre-test, post-test and delayed post-test design has become the standard approach to studies looking at incidental vocabulary learning (Schmitt 2010; Nation & Webb 2011). Third, research on incidental vocabulary learning through listening is an under-researched area in both the L1 and L2 contexts. L1 findings are somewhat contradictory; Robbins & Ehri’s (1994) results provide support; however, Brett, Rothlein & Hurley’s (1996) study found that vocabulary learning through listening to a story did occur when supported with teacher explanation of word meanings but did not occur without support. L2 studies have provided some indication that
L2 incidental vocabulary learning does occur through listening (Vidal 2003; Brown, Waring & Donkaewbua 2008; van Zeeland & Schmitt 2013b), but the amount of learning is less than might occur through reading (Brown, Waring & Donkaewbua 2008; Vidal 2011). Brown, Waring & Donkaewbua’s (2008) study is the only one that specifically examined incidental learning of L2 vocabulary through listening to stories. The effects of explaining word meaning during listening is yet to be examined in L2 research.

Given the importance of the study, the lack of research that has further examined incidental vocabulary learning through listening, and the fact that the experiments are clearly described, there is sufficient justification for replication. The study might be replicated in three ways. First, an approximate replication that included more detailed analysis of the results would be most useful. Additional analysis of the results would include using a paired-samples t-test in Experiment 1 to determine the statistical significance of any learning that occurred from pre-test to post, and determining the effect size. A mixed between-within participants repeated measures ANOVA should be included in the analysis of the second experiment to determine whether gains from pre-test to post-test are statistically different for each of the three groups, and whether the size of gains differs between the three groups. In the analyses of the gains between groups and between proficiency levels, relative gains should be examined rather than actual gains because relative gains account for differences in prior knowledge between groups (Shefelbine 1990; Horst, Cobb & Meara 1998). Approximate replication of the study with the additional analyses would help to substantiate the strength of the earlier findings.

Second, an approximate replication with L2 learners is needed to determine whether the original findings might also be valid in the L2 learning context. The procedure of the original study could be followed, along with the additional analysis suggested above. One variation that may be necessary in an L2 replication is the selection of texts. Research has indicated that children’s texts may be more lexically demanding than materials written for L2 learners (Webb & Macalister 2013). If the original texts are appropriate for advanced L2 learners, then the replication could be conducted accordingly. However, if the texts were found to be too difficult, the replication could be conducted using graded readers since these are the texts most commonly used in L2 extensive reading programmes. This would also increase the ecological validity of an L2 study. Graded readers would need to be carefully selected so that target words were chosen in relation to the six variables that Elley included in his study.

Third, an approximate replication of the study with L2 learners that involved additional manipulation of the number of readings of the text would provide greater clarity of how vocabulary is learned through listening to stories. In the original study, the stories were read to the learners three times on the basis of repeated listening of children’s stories being ecologically valid. However, it is not clear to what extent vocabulary may be learned from listening to stories once or twice with and without explanation of word meanings. Replication of the study with different matched groups that listen to the story once, twice and three times in each experiment would provide a more accurate indication of how listening and teacher explanation of word meanings contribute to vocabulary learning. It would show whether we might be able to generalize any gains in vocabulary knowledge to contexts where a story is heard once, or whether it is rather the accumulated gains made through repeated listening to the text that account for learning. Because L2 research has indicated that frequency plays a large role in whether or not words are learned incidentally through reading while...
listening (Brown, Waring & Donkaewbua 2008), and that repeated readings of a text increases the amount of vocabulary learning (Horst & Meara 1999), examining the degree to which words are learned through different numbers of readings would add greater clarity to Elley’s findings. Moreover, since discussion of vocabulary learning is typically related to reading or listening to a text once, this replication would help us to interpret the ecological validity of the original study.

4. Conclusion

Although vocabulary learning through meaning-focused input is widely researched, the two studies discussed in this paper represent areas within the field where further research is needed. Studies of lexical coverage have indicated the importance of vocabulary for comprehension of both written and spoken input (Schmitt, Jiang & Grabe 2011; van Zeeland & Schmitt 2013a), as well as the vocabulary sizes that may indicate comprehension of different discourse types (Meara 1991, 1993; Hirsh & Nation 1992; Adolphs & Schmitt 2003; Nation 2006; Webb & Rodgers 2009a, 2009b). Despite the interest that this line of research has generated, Liu & Nation’s (1985) study remains the only one that has looked at how lexical coverage affects guessing from context. Replication of this study would help to verify the earlier findings, allow us to generalize the results to other texts, and would thus be a useful starting point for further research in this area. The case for replication of Elley’s (1989) study is similar. It is the landmark study of vocabulary learning through listening and represents one of few attempts to research this area despite the importance of the topic to the field. Approximate replication of the study would help to verify the earlier findings, and determine whether the results could be generalized to the L2 context. Approximate replications could also then be designed to assess the robustness and generalizability of the findings. This approach to replication would help to determine the extent to which any vocabulary learning gains were due to listening to a story once, or whether they were the result of listening to multiple readings of the story. This would enhance our understanding of the findings because most discussion of incidental vocabulary learning through listening may be focused on listening to a text only once.

Acknowledgements

I am grateful to Paul Nation for his helpful comments on an earlier draft of this article.

References


van Zeeland, H. & N. Schmitt (2013a). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? Applied Linguistics 34.4, 457–479.


**Stuart Webb** is a Professor at Western University, Faculty of Education, Room 1026, Althouse Building, 1137 Western Road, London, Ontario, Canada. His research interests include vocabulary studies, extensive reading, and language learning using television. His articles have been published in journals such as *Applied Linguistics*, *TESOL Quarterly* and *Language Learning*. His recent book (with Paul Nation), *Researching and analyzing vocabulary*, was published by Heinle (2011).