UNIVERSITY OF LEEDS

Piloting a Doctoral Project on Vocabulary Learning: Procedure and Reflection

Xuechun Huang School of Education, University of Leeds

Abstract of the Article

Although pilot studies are important in vocabulary research, there is limited practical guidance on how to carry one out. It becomes more difficult when the pilot must be conducted outside the classroom in an unfamiliar location, involves face-to-face meetings with participants, and when the researcher has limited prior knowledge of the target group. This article addresses this gap by offering a researcher's perspective on the practical implementation of a pilot study on vocabulary learning, including methodological decisions, emergent challenges, and subsequent adaptations to the main study. This article first introduces the background and aims of the pilot study. Then, it describes the procedure of the pilot study, followed by reflections on its benefits and challenges. Finally, it highlights the significance of pilot studies for similar doctoral projects.

Introduction

Despite the large number of words that English as a foreign language (EFL) learners need to know for adequate communication, EFL learners' vocabulary growth rate is slow (Webb and Chang, 2012). To help them overcome these challenges, a major strand of vocabulary research is to explore the use of meaning-focused input (MFI) sources for EFL learners to learn vocabulary incidentally (e.g., Teng, 2024). Studies in this line found that learning can happen from exposure to these sources. However, the gains from MFI are small (Webb et al., 2023). To optimise vocabulary learning, several frameworks such as the Involvement Load Hypothesis (Hulstijn and Laufer, 2001) and the Technique Feature Analysis (Nation and Webb, 2011) have been proposed and studied. Generally speaking, these frameworks hypothesised that activities which were more motivating and cognitively demanding would lead to better vocabulary learning outcomes.

While these frameworks allowed researchers to evaluate vocabulary learning activities (e.g., flashcards, sentence writing), they did not provide guidance on how language lessons or courses should be organised as a whole. Nation's Four Strands Principle (2007) instead offers another potential way of learning from MFI from the perspective of course organisation. That is, combining MFI with the other three strands: meaning-focused output (MFO), language-focused learning (LFL), and fluency development (FD). However, previous studies have not examined the effects of one, two, three, and four strands on vocabulary learning. The aim of this doctoral research project is to compare learning gains of 10 target words between the following conditions: MFI only, MFI+LFL, MFI+LFL+MFO, and MFI+LFL+MFO+FD. This article reports on the procedure and reflection of conducting a pilot study.

Nation (2007) has not set a rule for the sequence of the Four Strands. However, according to the three-stage Skill Acquisition Theory (DeKeyser, 2020), the first stage of L2 learning is establishing declarative knowledge. This stage is relevant to MFI and

LFL, and having MFI before LFL may encourage learners to infer words from contexts on first exposure. The second stage is procedural knowledge. This stage is relevant to MFO, where learners try to use the words. The third stage is automatization. After proceduralization, much more efforts are needed before the knowledge can be accessed quickly, reliably, and easily, which is related to FD. Therefore, the sequence of the Four Strands in this study was MFI first, then LFL, followed by MFO and FD.

This project focuses on form-meaning links. The form-meaning link means that when learners encounter the words, they can think of the meaning, or vice versa. It is the most important as it creates the foundation for further development of other aspects of vocabulary knowledge (Schmitt, 2010a). I use the terms 'recognition' and 'recall' in describing vocabulary knowledge following Schmitt (2010b) because they also refer to test formats. Recognition means that learners can select the target words from several choices, while recall means that learners can provide the answers themselves.

Recent research shows that learning at recognition levels always takes place before recall levels (González-Fernández and Schmitt, 2020). This confirms Laufer and Goldstein (2004) that the following aspects of form-meaning links have different levels of learning difficulty (from easy to difficult): meaning recognition (choosing the correct meaning of the word from several options), form recognition (choosing the correct spelling of the word from several options), meaning recall (providing the meaning of the word), and form recall (providing the spelling of the word).

However, not all target words and test measures are appropriate for a particular study. If learners have fairly good knowledge of the target words, then a ceiling effect may occur as there is limited room for improvement in learning gains (Teng, 2016). The same may happen for test measures that are too easy for the participants. By contrast, if a test measure is too difficult, there can be a floor effect, where participants all score very low (Perez, 2020).

For this particular study, piloting was even more essential because I did not have my own classes, nor could I recruit participants from my workplace. Instead, I had to recruit participants widely from my professional network. In other words, without testing the participants directly, it would be difficult to gain a sufficient understanding of their prior knowledge of the 10 specific words. The first aim of the pilot study was to find out whether participants knew the target words already. Second, the pilot study aimed to evaluate which test formats are likely to best capture the learning gains: Chinese-to-English translation (form recall), checklist (form recognition), English-to-Chinese translation (meaning recall) or multiple-choice questions (meaning recognition).

Participants

The pilot study was conducted with 13 participants in a university in eastern China. They were recruited from opportunity sampling via English language teachers at that university. A recruitment poster was circulated in the group chats of their English language class, and students who were willing to participate contacted me. Ten of them were first-year undergraduates, and three of them were postgraduates. They were studying different majors, including artificial intelligence, software engineering, electronic science and technology, mechanical engineering, environmental science, electronic information, applied statistics, and French. They have learned English for 12.7 years on average. None of them had stayed in English-speaking countries for more than three months.

Intervention

The MF, LFL, MFO, and FD strands were operationalised as the following activity respectively: (1) reading while listening to a story of about 1,000 words, (2) seeing the flashcards of 10 target words, (3) writing to describe pictures that using the target words, (4) reading while listening to the story again. Participants were randomly assigned to an MFI only (n = 3), MFI+LFL (n = 2), MFI+LFL+MFO (n = 2), MFI+LFL+MFO+FD (n = 4), or a control group (n = 2). Randomising people into small groups, although unlikely to have much statistical power, can test the randomisation procedures, pilot the test link for each group (see the next paragraph), and all related materials. It can also estimate the time needed for each group to complete all activities. The pilot participants were similar to the main study participants in that they were also undergraduate students with about 12 years of experience learning English. Participants in both the pilot and main studies began learning English at primary school, as it became a compulsory subject at that level of education in China.

Data Collection

Data were collected from Qualtrics, a survey platform. The pilot study lasted for three weeks. Each week, participants received a Qualtrics link for the group they were assigned to, which included all the learning and testing activities. Weeks 1 and 3 were completed online with individual meetings with the participants. Week 2 was carried out in a face-to-face meeting with individual participants. The reason was that Week 2 was the week for intervention and was the most important session for the whole study. Meeting participants face-to-face helped me observe how participants engaged with the learning and testing activities.

In Week 1, the participants first completed an Updated Vocabulary Levels Test (uVLT) (Webb et al., 2017) and then the pre-test.

The uVLT measures knowledge of the most frequent 5,000 word families in English. Word families include base words and their inflected and derived forms. The purpose of uVLT was to inform the selection of target words. Target words usually need to come from a frequency band beyond the participants' current vocabulary levels. In Week 2, the participants completed their respective treatments and took the immediate post-test. After that, they provided feedback through a short, informal interview. In Week 3, the participants took the delayed post-test. The pre-test, immediate post-test, and delayed post-test had the same test items, but to minimise the test effect, the test items were reordered randomly.

Measures of Learning Outcomes

Form recall was measured through translation. Participants saw some Chinese sentences and their English translation. In the English sentences, one word was missing. Participants needed to fill in the blank. Participants saw one test item per page. In the pilot study, the initial letter of each target word was provided for the learners, but not the number of letters (e.g., a_____). This is to avoid participants giving synonyms as answers. However, it turned out that some participants still provided synonyms with the same initial letter (e.g., apparel instead of attire, brigand instead of bandit). Therefore, in the main study, participants saw both the initial letter and the number of letters of the answers (e.g., a _ _ _ _ _).

The parts to measure form recognition and meaning recall contained 34 items, with 10 target words, 21 distractors, and three words participants should be familiar with. The distractors were pseudowords. Participants saw a word and were provided with the pronunciation of the word. They needed to choose between "I know this word" and "I don't know this word". Participants saw one test item per page. Then, on the next page, they were asked to provide anything they knew about the meaning of the word if they had chosen "I know this word". They were told to leave the box blank and move to the next page if they had chosen "I don't know this word". Distractors were revised if they looked too similar to the participants' known words or if they often tempted the participants to guess the meaning based on the form. For example, the distractor 'rootee' was chosen as many as nine times across the three weeks, and participants who chose it consistently suggested that it meant 'related to root'. Therefore, this distractor was changed to 'kootee'.

A multiple-choice test measured participants' meaning recognition of the words. Participants were asked to choose only one option from six. The six options were a key, four distractors and one "I don't know" option. However, it was found that the multiple-choice test was too easy for the participants; they could score very high in the pre-test, so the amount of learning gains that could be captured from pre- to post-tests was much smaller compared to other measures. In addition, I asked one of their teachers which test formats would be appropriate to measure the learning gains of my target words for these groups of learners. She told me that form recall could be difficult, and meaning recall seemed the most appropriate. Such conversations with teachers also deepened my understanding of participants.

Ethics

This study was approved by the Faculty Research Ethics Committee for Business, Environment, and Social Sciences at the University of Leeds (Approval Number: 1689). All participants read an information sheet about the study and then signed a

written consent form before voluntarily taking part in the study. Their data were anonymous. Their names were replaced by a number (e.g., ID = 1). Participants were informed that they could withdraw from the study before the second week.

Reflection

The pilot study, which was a rehearsal of the whole data collection procedure, first made me aware of the difficulty of recruiting participants. Due to busy schedules, many teachers said that it was not feasible to use regular class hours for their classes to take part in the pilot study. However, they were happy to circulate the recruitment poster in their classes to enable students to take part outside the classroom. In the main study, to increase the sample size, I communicated with the teachers to find the most convenient time, which allowed me to recruit several classes to complete the study during normal class hours. Moreover, the pilot study allowed me to estimate the time that participants needed to complete each week. I could therefore communicate with the gatekeepers and teachers precisely how much time I needed in the class for the main study. I was also able to remind participants in the main study to speed up if I found they spent too much time in some parts of the test.

One challenge was that the pilot study itself was intensive. I had to arrange individual meetings with each of the 13 participants every week for three consecutive weeks, starting as early as possible to avoid clashing with their exam weeks. As I was not a member of the university, I could not book a meeting room. Instead, I met them either online (Weeks 1 and 3) or in a public café on campus (Week 2). I needed to access the campus as a visitor and, most importantly, decide where to meet my participants. I first identified all the cafés on campus through the campus map I found online. I searched online for posts where people had shared which cafés were quiet, had sockets, and were open to the public. I then scheduled the meeting time with the participants according to an ideal café's opening hours and shared the address with them. Another challenge was that the sample size was small, so that each participant's in-depth responses were crucial. Initially, I was concerned about the amount of information I could gather. However, all participants were highly engaged and provided valuable insights. The participants also generously offered feedback for the learning materials, which greatly improved the quality of the study.

The pilot study contributed to confirming and refining the research methods. The pilot study first aimed to find out to what extent the target participants knew the target words at the form recall, form recognition, meaning recall, and meaning recognition levels. The results showed no word was known to over 80% of the participants at any level. Therefore, all target words were kept in the main study.

The pilot study also aimed to identify which test formats would best capture the learning gains. The results indicated that three test measures (i.e., form recall, form recognition, and meaning recall) could potentially distinguish between the experimental groups and the control group, with the experimental groups having much higher learning gains. Therefore, these measures were kept in the main study. The meaning recognition test was removed in the main study for two reasons. First, having too many test measures would take a long time and lead to test fatigue. Second, it showed a ceiling effect, as participants could perform near the maximum score at pre-test, leaving little space to demonstrate measurable improvement.

However, all the revisions would not have been possible if I had not planned the time to adjust between the pilot and main studies. The time needed for such adjustments may vary depending on the nature of the project. I had about two months to examine the results of the pilot study, reflect on the entire process, and make changes accordingly.

Conclusion

This article has discussed the procedure and reflection of a pilot study on vocabulary learning. The pilot study helped examine the feasibility of the intervention, confirm target words and materials, estimate the time needed for each week, and evaluate instruments. The insights obtained from the pilot study were critically incorporated into the main study.

This article highlights that requesting class hours from teachers for both pilot and main studies can be very challenging; instead, recruiting individual students to complete the pilot study outside the classroom may be more acceptable. Conducting pilot studies in unfamiliar sites is indeed difficult but feasible with careful planning. This article encourages the vocabulary research community to explore alternative, practical approaches to conduct pilot studies in order to increase the pool of participants and overcome the challenge of restricted classroom access.

References

DeKeyser, R. 2020. Skill Acquisition Theory. In: VanPatten, B., Keating, G.D. and Wulff, S. eds. *Theories in Second Language Acquisition.* Third ed. New York: Routledge, pp.83–104.

González-Fernández, B. and Schmitt, N. 2020. Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics.* 41(4), pp.481-505.

Hulstijn, J.H. and Laufer, B. 2001. Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition. *Language Learning.* 51(3), pp.539-558.

Laufer, B. and Goldstein, Z. 2004. Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning.* 54(3), pp.399-436.

Nation, I.S.P. 2007. The Four Strands. *Innovation in Language Learning and Teaching*. 1(1), pp.2-13.

Nation, I.S.P. and Webb, S. 2011. *Researching and analyzing vocabulary.* First ed. Boston, MA: Heinle, Cengage Learning.

Perez, M.M. 2020. Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition.* 42(4), pp.749-773.

Schmitt, N. 2010a. Key issues in teaching and learning vocabulary. In: Chacón-Beltrán, R., Abello-Contesse, C. and Torreblanca-López, M.d.M. eds. *Insights into Non-native Vocabulary Teaching and Learning. Multilingual Matters*, pp.28-40.

Schmitt, N. 2010b. *Researching vocabulary: a vocabulary research manual.* Basingstoke, Hampshire: Palgrave Macmillan.

Teng, F. 2016. The effects of word exposure frequency on incidental learning of the depth of vocabulary knowledge. GEMA Online® *Journal of Language Studies.* 16(3), pp.53-70.

Teng, F. 2024. Working memory and prior vocabulary knowledge in incidental vocabulary learning from listening, reading,

reading-while-listening, and viewing captioned videos. *System*. p.124.

Webb, S. and Chang, A. 2012. Second language vocabulary growth. *RELC Journal.* 43(1), pp.113-126.

Webb, S., Sasao, Y. and Ballance, O. 2017. The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *International Journal of Applied Linguistics.* 168(1), pp.33-69.

Webb, S., Uchihara, T. and Yanagisawa, A. 2023. How effective is second language incidental vocabulary learning? A meta-analysis. *Language Teaching.* 56(2), pp.161-180.