Learning vocabulary through listening: the role of vocabulary knowledge and listening proficiency

Article

Published Version

Open Access

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1111/lang.12411

Publisher: Wiley

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

EMPIRICAL STUDY

Learning Vocabulary Through Listening: The Role of Vocabulary Knowledge and Listening Proficiency⬢

Pengchong Zhang (iD) and Suzanne Graham (iD)

Institute of Education, University of Reading

This study explored the impact of preexisting vocabulary knowledge (PVK) and listening proficiency on the vocabulary learning through listening of 137 Chinese learners of English, when provided with three types of oral vocabulary explanations—second language (L2), codeswitching (CS), and contrastive focus-on-form (CFoF)—and when no explanations (NE) were provided (extending Zhang & Graham, 2019). Listening proficiency was a more important factor influencing vocabulary learning through aural input than PVK was, with most notable gains for learners with high listening proficiency and low PVK. The CFoF approach was the most helpful for learners regardless of their PVK and listening proficiency, whereas the NE approach was the least helpful. Moreover, comparing just the CS and L2 groups, the CS approach was more helpful for lower PVK learners and for more proficient listeners than the L2 approach was. Higher PVK learners and less proficient listeners, however, benefited more from the L2 approach than from the CS approach. The study highlights the complex interplay of vocabulary

The study highlights the complex interplay of vocabulary

knowledge, listening proficiency, and instructional conditions, factors useful to bear in mind when planning activities to enhance vocabulary learning through listening.

Keywords vocabulary learning; proficiency; listening; instruction; focus on form

Introduction

Helping learners to gain a wide range of vocabulary knowledge is a fundamental issue for improving their general language proficiency globally, and no less so in classes teaching English as a foreign language (EFL) in China (Silver, Hu, & Iino, 2002). Discussion often focuses on the comparative merits of intentional learning and of incidental learning, where learners unintentionally "pick up" vocabulary knowledge when they are focused on understanding the meaning of the language input (Hulstijn, 2001). Whereas vocabulary gains tend to be smaller for the latter (Laufer & Girsai, 2008), perhaps because learners' attention is on global meaning rather than on individual vocabulary items, pedagogical activities can be used alongside a focus on meaning to enhance the salience of items (Sharwood Smith, 1991) and hence noticing (Schmidt, 1990) and thence vocabulary learning (Laufer, 2005).

The extent to which learners of different proficiency levels in a second language (L2) benefit from different types of attention-enhancing pedagogical activities is, however, underexplored, particularly in relation to aural input and the role of the first language (L1). A consideration of those issues is important, not only from a pedagogical perspective, but also within models of L2 vocabulary acquisition in which the L1 functions differently at different levels of proficiency (Jiang, 2002, 2004; Kroll & Stewart, 1994). This study therefore investigated the impact of L2 learners' preexisting vocabulary knowledge (PVK) and listening proficiency on their vocabulary learning from attention-enhancing activities in the form of different types of vocabulary explanations after listening: no explanations (NE), explanations in the L2, codeswitched explanations, and those that give contrastive focus-on-form (CFoF) information.

Background Literature

The extent to which vocabulary knowledge can be acquired through listening has received less research attention than is the case for reading. A common thread running through the literature, however, is, first, that levels of vocabulary learning from listening are typically much lower than for reading (Brown, Waring, & Donkaewbua, 2008; Vidal, 2011); and, second, that listening may develop what van Zeeland and Schmitt (2013a) call the earlier-acquired aspects of vocabulary knowledge, such as form recognition, rather than "high levels

of knowledge" (p. 609), such as meaning recognition and recall. Amounts of learning through listening may also depend on learners' level of proficiency, in the form of either general proficiency, prior vocabulary knowledge, or specifically listening proficiency. For example, Vidal (2011), in a study comparing incidental vocabulary learning by university-level L2 learners through both reading and listening, found that for pretest to posttest gains, the higher learners' general proficiency level was, the smaller was the difference between gains made from reading and gains made from listening. Lower proficiency learners, especially at the very lowest levels of proficiency, retained very few words from the spoken input (fewer than from reading), possibly, according to Vidal, because they were hampered by difficulties in segmenting words from the speech stream, rendering attention to vocabulary problematic.

Existing vocabulary knowledge specifically might also be expected to influence how much vocabulary is gained through spoken input, given that vocabulary breadth is positively correlated with listening proficiency (Stæhr, 2008); learners with larger vocabulary sizes might go on to acquire more words from aural input because they comprehend more of the input in the first place. This has been found to be the case for studies of incidental learning through listening as part of video viewing (Peters & Webb, 2018). By contrast, Rodgers (2013) found no effect of vocabulary knowledge on learning gains from television viewing among intermediate university learners of English.

These somewhat contradictory findings may result from the more complex relationship between vocabulary knowledge and listening comprehension compared with vocabulary knowledge and reading, with implications for how important vocabulary knowledge is for vocabulary learning through listening. First, a wide range of correlations between vocabulary knowledge and listening has been found in previous research, varying between $r = 0.209$ to $r = 0.941$ in an unpublished meta-analysis of 26 studies between 2000 and 2018 (Smith, 2019). This range may reflect differences across studies in how vocabulary knowledge has been measured. Commentators such as Stæhr (2009) argue that aural vocabulary tests should be used rather than written tests in explorations of the relationship, a view supported empirically by Cheng and Matthews (2018).

Second, even though some strong correlations have been reported between vocabulary size and listening comprehension, generally speaking they tend to be lower than those reported for reading (Stæhr, 2008). Vocabulary knowledge explains less variance in listening compared with other skills (Miralpeix & Muñoz, 2018), and lower levels of vocabulary knowledge seem to be needed for aural comprehension than for reading comprehension (van Zeeland & Schmitt, 2013b). L2 listeners can, and may indeed have to, draw on factors other than

vocabulary knowledge when interpreting spoken input, for example, contextual information from gesture, tone of voice, and facial expression. Thus higher vocabulary knowledge does not always equate to better listening; for example, van Zeeland and Schmitt (2013b), in a study of university-level learners of English, found that higher levels of vocabulary knowledge generally led to higher levels of comprehension, but that there were also learners with lower levels of lexical knowledge but "adequate" comprehension. Similar findings are reported by Bonk (2000) for Japanese university learners of English, some of whom with lower levels of lexical knowledge achieved good comprehension, whereas others with higher levels had quite poor comprehension. These studies suggest that it might be useful to consider the effect of listening proficiency as well as vocabulary knowledge on how much vocabulary is acquired through spoken input.

The extent to which vocabulary knowledge is developed through listening may also depend on the type of listening activity engaged in and the types of additional support offered. Vocabulary learning during listening for meaning can be enhanced through a lexical focus-on-form approach (Laufer, 2005; Laufer & Girsai, 2008), in which the salience of lexical items is heightened so that the learners' attention is drawn in some way to phonological, orthographic, and semantic information about vocabulary items, thus increasing noticing (Schmidt, 1990). Indeed, the fleeting nature of spoken input means that lexical focus-on-form through some kind of input enhancement may be especially important for vocabulary learning through listening, where "focal" attention (Ellis, 1999, p. 35) is typically on broad understanding and only "peripheral" attention is given to individual linguistic items (Vidal, 2011). Enhancement of spoken input may include among other things the provision of video captions or annotations (Montero Perez & Desmet, 2012), or explanations or elaborations of the target items (Hennebry, Rogers, Macaro, & Murphy, 2013; Tian & Macaro, 2012; Vidal, 2011).

Different forms of input enhancement are likely to vary in the degrees of "involvement load" (Hulstijn & Laufer, 2001, p. 539) they prompt, however, and hence potentially lead to different amounts of learning. Involvement load is influenced by "need," "search," and "evaluation" (p. 539): that is, to what extent learners have to understand a given item for task completion (need), seek out the item's meaning (search), or consider whether a given meaning or use for a word is the most appropriate one for a certain context (evaluation). The level of evaluation is considered to be especially important (Hulstijn & Laufer, 2001), as more evaluation usually brings with it deeper processing and therefore is more likely to lead to better learning outcomes.

Laufer and Girsai (2008) draw on such a framework to explain greater gains made by learners experiencing a lexical focus-on-form approach in the context of reading. This approach, which they termed "contrastive form-focused instruction," provided input on "the similarities and differences between [the] L1 and L2 in terms of individual words and the overall lexical system" (p. 696). Further evidence of the potential value of contrastive form-focused instruction as a type of vocabulary enhancement emerged from a study by Zhang and Graham (2019) in the context of listening, an investigation whose data we returned to for the present study. School-aged learners in China listened to a series of passages in English and then received one of four types of instruction: L2 explanations of vocabulary items in the passages; codeswitched (L1) explanations; explanations providing additional crosslinguistic information (CFoF); or NE but rather cultural information related to the passages. For short- and long-term vocabulary learning, the three treatment groups significantly outperformed the NE group. Whereas no statistically significant differences were found between the L2 and codeswitching (CS) groups for short-term and long-term improvement, gains for the CFoF group were significantly greater than for the L2 and CS groups (all as reported in Zhang & Graham, 2019). As the focus of the study was on exploring whether different types of input enhancement had differing impacts on vocabulary learning through listening, learners' PVK and listening proficiency levels were not of primary interest in that study, but rather were treated as covariates (i.e., two continuous control variables) in the data analysis to control any baseline differences between the intact groups included in the quasi-experimental design.

The study by Zhang and Graham (2019) is one of a growing number exploring vocabulary enhancements in the form of explanations by the teacher, either before or after listening. Such explanations are likely to be especially important for lower proficiency, beginning language learners in a classroom setting, where listening activities are often carried out under the guidance of the teacher rather than independently. For example, Pujadas and Muñoz (2019) conducted a 1-year intervention with 74 secondary school learners of English in Grade 8 (aged 13–14 years) in Spain. Learners watched 24 episodes of a TV series in one of four conditions: (a) L2 captions (written on the screen) and vocabulary preteaching, (b) L2 captions and no preteaching, (c) L1 subtitles and preteaching, and (d) L1 subtitles and no preteaching. All groups made significant vocabulary gains, but preteaching of items, with either L2 captions or L1 subtitles, led to the greatest gains. Higher general proficiency (measured by Oxford Placement Test scores) was related to higher vocabulary learning

gains, with learners at the A2/B1 level showing significantly greater gains than those at A1 or Pre-A1.

A slightly larger group of studies has considered the impact of post-listening explanations on vocabulary learning, also with a focus on the interaction between learner proficiency level and types of explanations. Working with university EFL learners within a lexical focus-on-form approach, Tian and Macaro (2012) explored the impact of post-listening explanations that used either the L1 (CS group) or the L2-only. For analysis, learners were allocated to four proficiency levels according to their scores on a listening test and a vocabulary pretest. The authors hypothesized that lower proficiency learners would benefit more from L1 than from L2 explanations. This would be supported by models of vocabulary acquisition such as those of Jiang (2002, 2004) and Kroll and Stewart (1994), in which the L1 is the dominant language at the initial stage of L2 learning, and L2 words are lexically mediated by the L1. In Tian and Macaro's study, the CS group outperformed the L2-only group on short-term vocabulary learning, but the interaction between group and proficiency levels was not statistically significant. That is, learners in both treatment groups made significant pre–post vocabulary gains regardless of their proficiency level.

Tian and Macaro (2012) suggested that the absence of a proficiency effect may be attributable to the low frequency level of the target items in their study, for which the higher proficiency learners as well as the lower proficiency learners needed support from the L1. An absence of proficiency effect was also reported in a subsequent study with a similar design but with younger, high school learners of French (Hennebry et al., 2013). Different findings emerged, however, from a study closely modeled on that of Tian and Macaro among adult EFL learners, where Lee and Levine (2020) investigated whether learners' proficiency level (intermediate vs. advanced) interacted with the two types of vocabulary instruction provided for listening input (L2 English-only vs. codeswitched L1 vocabulary explanations). Results indicated that whereas advanced learners learned and retained similar amounts of vocabulary knowledge across the two teaching conditions, the intermediate-level learners who received codeswitched L1 explanations significantly outperformed those who were given L2 English-only explanations in terms of long-term vocabulary retention. Additionally, they gained as much vocabulary knowledge as the advanced learners within the codeswitched L1 group.

Important though these studies are for their insights into vocabulary learning through listening, it should be noted that all used proficiency level as a

categorical rather than an interval variable, making the analyses less sensitive than they might have been. This is because categorizing continuous predictors into groups (by using, for example, a median split) may suffer from some limitations (Aiken & West, 1991). First, the scale of the values within each category is highly likely to be skewed. Second, two very close values can be arbitrarily allocated into the lower band of a "high" category and the higher band of a "low" category. Moreover, by dividing a sample in this way, all values falling into one category are considered to be equivalent, when in fact they are not. In sum, the variance offered by a continuous variable is lost when it is converted into a categorical data, and it is such variance in proficiency that may reveal its potentially subtle effects.

Researchers thus have relatively few clear insights into the role of proficiency level in vocabulary learning through spoken input, and in particular in relation to the interaction between proficiency and vocabulary enhancement type, with contradictory findings within the small number of studies that have been undertaken. Additionally, relatively little attention has been paid to these issues in school settings, where questions of proficiency and how it interacts with instructional approaches are especially important. The current study therefore addressed the need for further research of this kind.

The Current Study

In the current study we returned to our previous study of high school vocabulary learning through listening (Zhang & Graham, 2019). We used data collected for that study but adopted a different perspective on the learning differences between the CS, L2, CFoF, and no explanation (NE) groups by going one step further: We included learners' PVK and listening proficiency as two predictors (of primary interest within the study design) and, more importantly, explored the interaction between these two predictors and other predictors, namely, (a) the time points at which the tests for the target vocabulary items were administered and (b) the four types of vocabulary explanation.

We posed two research questions in relation to post-listening vocabulary explanations (for L2, CS, CFoF, and NE conditions):

1. To what extent does vocabulary learning through listening, taking all conditions together, vary according to learners' PVK and listening proficiency?
2. To what extent does the impact of the different types of post-listening vocabulary explanation vary according to learners' PVK or listening proficiency?

Method

Sampling and Procedures

Following Tian and Macaro (2012), the study employed a quasi-experimental design, involving 137 first-year senior-secondary school EFL learners from four intact classes in China from one school (aged 15–16, with 7 years' English learning experience). Learners were preparing for the Gaokao, China's national university entrance exam, and hence had a proficiency level of around A2 to B1 on the CEFR, the Common European Framework of Reference for Languages (Council of Europe, 2001), or around levels 3–4 on IELTS, the International English Language Testing System (https://takeielts.britishcouncil.org/teach-ielts/test-information/scores-explained). Classes were randomly assigned to three treatment groups—a second-language (L2) group ($n = 35$), a teacher-CS group ($n = 36$), and a CFoF group ($n = 33$)—and one no-explanation (NE) group ($n = 33$). All groups were instructed by the first author (whose L1 is Chinese). In our choice of group names, we followed Tian and Macaro (2012) in the use of "codeswitching" to signal "principled rather than ad hoc L1 use" by the teacher (p. 69), although we acknowledge, as they did, that the term has a somewhat different meaning outside of the classroom. The essence of the CFoF approach was providing crosslinguistic information about items' use rather than simply giving their meaning in the L1 (Laufer & Girsai, 2008).

The data collection procedure began with a general vocabulary knowledge test (GEVT), a vocabulary pretest, and a listening comprehension test (week 1). Six teaching sessions took place between weeks 4 and 9. All groups completed a vocabulary posttest at the end of each session. From the third teaching session inclusive (week 6), an additional vocabulary delayed posttest was administered at the same time as the vocabulary posttest. There were six delayed posttests in total, with the final two administered in weeks 10 and 11 after the completion of the teaching sessions. Each delayed posttest assessed long-term vocabulary retention for target items from the session delivered two weeks previously. We ensured that all target items received an equal delay of two weeks between the posttest and the delayed posttest for those items. Timings for all aspects of individual lessons (including test administration) were tightly controlled as part of detailed lesson plans strictly followed by the teacher, not only to ensure uniformity across the groups but also to adhere to the 45-minute lesson time required by the school. The study design is outlined in Figure 1.

Intervention Procedures and Materials

The intervention was implemented over six sessions for all groups. In each session, learners heard a prerecorded listening passage (once) and then answered
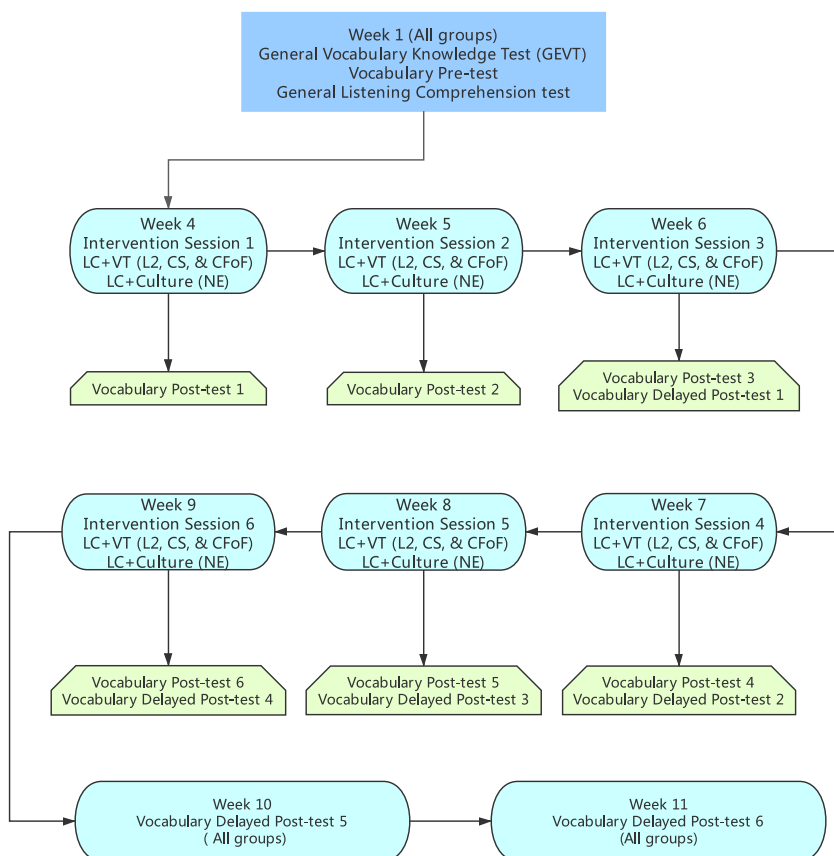
Figure 1 Study design. LC = Listening Comprehension Test; VT = Vocabulary Test; L2 = second language; CS = codeswitching; CFoF = contrastive focus-on-form; NE = no explanations. [Color figure can be viewed at wileyonlinelibrary.com]

three written comprehension questions, one eliciting understanding of the global meaning and two eliciting their understanding of specific and important details, so that learners' focus was on comprehending what the passages were about. The nature of the intervention then differed between the NE group and the three treatment groups (L2, CS, and CFoF). For the latter groups, the listening passage was played once more, sentence by sentence. More specific comprehension questions were asked orally at this stage in order to focus learners' attention on listening comprehension and to initiate active classroom participation. Subsequently, the teacher gave explanations of the

target vocabulary items, again geared toward meaning comprehension of the passage, but each of the three treatment groups received a different form of vocabulary explanation: L2 only, codeswitched, or CFoF.

Steps were taken to ensure that each treatment group received the same amount of vocabulary explanation for a specific vocabulary item (see Appendix S1 in the Supporting Information online for examples of explanations). Thus, the L2 group learners first received a short sentence in English (the L2) explaining the meaning of the target vocabulary item. Then, they were given an additional L2 sentence including the target item and were required to use the L2 explanation of the target item that they had been given in order to paraphrase in the L2 the meaning of the additional L2 sentence. In the CS group, the meaning of the target lexical item was given by the teacher in Chinese. Learners were also given an additional L2 sentence including the target item and were asked to show understanding of the meaning of the sentence by translating the sentence into Chinese. In the CFoF group, learners were initially given the L1 meaning of the target vocabulary item, and then an additional explanation was provided in the L2, focusing on comparing and contrasting the L2 word and its L1 translation, drawing attention to any mismatch between the two. All groups also saw the written forms of the target items, presented through PowerPoint, but they were not allowed to write them down.

Finally, the instructor read out the whole listening passage once more and repeated L2, CS, and CFoF explanations for the target items, stopping after each target item while reading out the text. All groups heard the passage three times in total. Learners would already have been familiar to a certain extent with all three approaches to vocabulary explanation, because these were normally adopted by their class teachers when explaining new vocabulary.

Learners in the NE group first heard the same listening passage once and completed the same written listening comprehension questions as learners in the three treatment groups. The listening passage was then replayed sentence by sentence twice for them. In addition, the teacher gave them culture and background information in the L2 relating to the listening passage but unrelated to the target vocabulary items. Therefore, like learners in the three treatment groups, learners in the NE group also heard the listening passage three times in total.

Instruments
Baseline Vocabulary Knowledge and Listening Comprehension Tests
Before the intervention, learners' PVK was assessed through an aural 160-item general vocabulary knowledge test (GEVT). This was based on the aural

1.
    A. 金钱 (money)
    B. 食物 (food)
    C. 时间 (time)
    D. 朋友 (friends)

Participants hear: Time, they have a lot of time.

Figure 2 Example item for GEVT (general vocabulary knowledge test).

vocabulary levels test (McLean, Kramer, & Beglar, 2015) and designed to measure learners' existing vocabulary levels and also their existing knowledge of the target vocabulary items for the intervention. Piloting indicated that the first three most frequent bands of 1,000 words (1K, 2K, and 3K) and the academic word list were appropriate for the proficiency level of the participants, thus our GEVT drew on 100 items from those lists. An additional 60 target items, measuring learners' partial knowledge of the target lexical items to be presented in the intervention, were intermingled with the 100 items, so that the participants would not know which items were the focus of the study. This made 160 items in total. See https://www.iris-database.org/iris/app/home/detail?id=york:937834 for the test. The test format, meaning recognition with multiple choice, allowed for the assessment of partial knowledge of a large number of items. The researcher read out the target lexical item, first on its own, and then in a sentence that gave no clue to its meaning. Participants then had to select the correct Chinese translation for the English word from four options (one correct Chinese translation and three distractors). An example item for this test is given in Figure 2. The English translations in parentheses are given for clarification and did not appear in the test paper. Cronbach's alpha for the test was .76.

Learners' listening proficiency was assessed at baseline through the first two sections of a standard IELTS listening test (https://takeielts.britishcouncil.org/take-ielts/prepare/free-ielts-practice-tests/listening. These tests cannot be made openly available as they are proprietary). These two sections were deemed appropriate for the level of the participants, because they were set in an everyday social context and hence assessed listening at around IELTS levels 3–4 (British Council, n.d.). Cronbach's alpha for this listening test was .62, comparable with what has been reported for subsections of the IELTS test elsewhere (Breeze & Miller, 2012).

Listening Passages

Materials for the intervention were chosen from an English textbook, *New Senior English for China* (Liu et al., 2007), in order to maximize the ecological validity of the materials used. Although aimed at learners of the same proficiency level as participants, the textbook was not used in the province where the school was located. Therefore, there was very little possibility of the participants having access to it before the intervention or outside of class. Six passages were identified and altered to ensure that they were on topics relevant to learners and were all around 250 words long. They were then turned into audio recordings by L1 English speakers, ensuring a speech rate of approximately 150–190 words per minute. That is, they were at the lower end of average speech rates for radio monologues and conversations in British English, according to Tauroza and Allison (1990).

The textbook had already highlighted the words and collocations that, based on the senior high school English curriculum (Ministry of Education, 2003), should be new and therefore taught to EFL students at the proficiency level of the participants in the study. The six listening passages were examined carefully, making sure that no more than 5% of the words in the passage were highlighted as new to learners, therefore reaching the threshold of 95% lexical coverage for listening comprehension (van Zeeland & Schmitt, 2013b). Two online vocabulary profilers (VP–Classic and VP–Compleat; Cobb, n.d.) showed that the items to be taught to learners consisted of (a) 43 single words, mainly from the 1K, 2K, and 3K most frequent bands and from the academic word list (18 nouns, 13 verbs, and 12 adjectives); and (b) 17 collocations, which we define as groups of words "that belong together, either because they commonly occur together . . . or because the meaning of the group is not obvious from the meaning of the parts" (Nation, 2001, p. 317). We therefore had 60 target items in total, 10 in each listening passage. All listening passages, their vocabulary profiles, and their target lexical items are provided in the Supporting Information online (in Appendices S2, S3, and S4, respectively). Learners in all four groups listened to all six passages.

As part of the teaching procedures, learners handed in their responses to the three initial, written comprehension questions, from each session, administered after the first hearing of each passage. Although they did not form part of the data used to judge the effectiveness of the interventions, the responses were reviewed to check that learners in each group had good comprehension of the six listening passages. On average, learners gave correct answers to more than two thirds of the 18 questions, and this was true across all four groups (L2: $M = 12.71$, $SD = 1.87$; CS: $M = 12.27$,

0 _____   1 2 3 4 5

Participants hear: Overcome, we need to overcome this.

Figure 3 Example item for the vocabulary posttest and delayed posttest.

$SD = 1.62$; CFoF: $M = 12.37$, $SD = 2.06$; NE: $M = 12.30$, $SD = 2.00$). In addition, a one-way ANOVA test indicated that the four groups were not statistically different in their comprehension of the passages, $F(2, 148) = 0.40, p = .75$, $\eta^2 = .01$.

Vocabulary Posttests and Delayed Posttests
The impact of the listening sessions on learners' vocabulary knowledge was assessed through a vocabulary posttest and delayed posttest based on the test used by Tian (2011) but modified so that it took an aural form in which the teacher read out one target vocabulary item plus an additional sentence including the item. See https://www.iris-database.org/iris/app/home/detail?id=york:937834 for the tests. Learners then had 10 seconds to give their response. If they did not know the meaning of the item, they were asked to circle 0 on the answer sheet. If they knew the meaning of the item, however, they were required to write it down either in Chinese or the very same word in English and to circle a number from 1 to 5 indicating the degree to which they felt confident about the meaning they provided. Figure 3 shows an example item. Hence, although meaning recognition was assessed at pretest, meaning recall (without multiple choice responses) was assessed at posttest and delayed posttest, to provide a more stringent measure of vocabulary growth (see the sections below, Data Analysis and Limitations and Future Direction).

Data Analysis
Vocabulary pretests, posttests, and delayed posttests were scored by the first author. Posttests and delayed posttests were then second-marked by another researcher, giving interrater reliability rates of between 98.76% and 99.22% (for further details of the scoring system, see Appendix S5 in the Supporting Information online). Items were marked either right or wrong, with no half marks, in order to permit the use of binary logistic regression tests.[1] An exploratory factor analysis was conducted with the six posttests and with the six delayed posttests (separately). Results indicated that the six posttests loaded onto one factor, explaining 77.71% of the variance. The six delayed posttests

also loaded onto one factor, explaining 57.56% of the variance. As these factor loadings were relatively high (.58 to .94), it was considered justifiable to aggregate scores for all six posttests and, separately, all six delayed posttests, giving one total score for each. The reliability for the aggregated vocabulary posttest was .94, and .92 for the delayed posttest (Cronbach's alpha). The GEVT was used to give an indication of participants' relative vocabulary level through the number of correct items out of a total possible score of 100 (i.e., excluding the items used in the intervention), rather than an estimate of their vocabulary size.

The quantitative data were analyzed both by participant (137 participants) and by item (60 vocabulary items, coded 1 if correct and 0 if wrong). Therefore, the outcome variable in our analyses was at a binary level. Binary logistic regression tests are normally recommended for analyzing data with a single binary outcome variable and one or more continuous or categorical predictor variables. In order to control both by-participant and by-item random effects, however, we decided to adopt generalized linear mixed effects models, which allowed us to run binary logistic regression tests with a random effects structure. The models were run with the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017), a package based on lme4, in R (version 3.5.0; R Development Core Team, 2018). Random effects were fitted using a maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013). In cases where a full random effects structure model did not converge, we first took out the interactions between random slopes, and then gradually removed random slopes that accounted for the least variance until a converged model was obtained.

Results

We first calculated descriptive statistics for the two baseline tests and the vocabulary pretests, posttests, and delayed posttests for each group, in order to explore (a) the extent to which vocabulary learning across all conditions together varied according to learners' PVK and listening proficiency, and (b) the extent to which the impact of different post-listening vocabulary explanations in the four groups varied according to learners' PVK or listening proficiency. Table 1 shows the results.

Four fixed effects were entered into our first model: Time (1. Pretest, 2. Posttest, 3. Delayed posttest); Group (CFoF, CS, L2, NE); GEVT (vocabulary test); Listening (listening comprehension test). Time 1 was set as the baseline for Time and CFoF was set as the baseline for Group. Three-way interactions involving the fixed effect of Time were also added to the fixed effects structure. The Time × GEVT × Listening interaction was examined to address the first research question, and the Time × Group × GEVT and Time × Group ×

Table 1 Descriptive statistics for the baseline tests and the vocabulary tests

Group (n)	Statistic	GEVT	Listening	Pretest	Posttest	Delayed posttest
L2 (35)	Mean (SD)	50.43 (8.06)	21.43 (11.35)	10.29 (3.30)	27.80 (8.46)	16.57 (8.00)
	Min–max	31.00–66.00	5.00–45.00	4.00–16.00	7.00–46.00	5.00–37.00
CS (36)	Mean (SD)	50.31 (6.60)	17.64 (10.18)	11.64 (4.85)	44.22 (7.48)	16.53 (8.21)
	Min–max	35.00–66.00	0.00–45.00	2.00–21.00	30.00–56.00	6.00–36.00
CFoF (33)	Mean (SD)	49.18 (8.69)	18.97 (12.63)	6.00 (1.44)	50.00 (6.64)	28.70 (9.16)
	Min–max	29.00–72.00	0.00–55.00	3.00–10.00	29.00–59.00	14.00–49.00
NE (33)	Mean (SD)	46.94 (7.69)	16.67 (11.43)	5.58 (1.73)	9.88 (5.72)	1.67 (2.04)
	Min–max	28.00–61.00	0.00–40.00	2.00–9.00	1.00–23.00	0.00–9.00

Note. The maximum score for the GEVT and listening test is 100.00; that for the pretest, posttest, and delayed posttest is 60.00. GEVT = general vocabulary knowledge test; L2 = second language; CS = codeswitching; CFoF = contrastive focus-on-form; NE = no explanations.

Listening interactions were examined to address the second research question. Because interactions between categorical and continuous predictors were included in the model and the two continuous predictors, GEVT and Listening, were not on the same measurement scale as each other, we entered the standardized z scores of these two continuous predictors into the model. In this way, the two continuous predictors were both mean centered and were on the same measurement scale.

The random factors included Participants and Items. The random effects structure of the converged model included intercepts for Participants and Items and by-Item random slopes for Time and Group. This model represented a good fit to the data, $R^2_{marginal} = .41$ (variance explained by the fixed effects), $R^2_{conditional} = .77$ (variance explained by both the fixed effects and random effects), and there was no significant overdispersion or collinearity (all generalized variance inflation factors [GVIFs] < 3.5). Table 2 shows the results. There were significant three-way interactions for Time $\times$ Group $\times$ GEVT (lines 32 and 33 in the table), Time $\times$ Group $\times$ Listening (lines 27 and 30), and Time $\times$ GEVT $\times$ Listening (line 38).

We did not, however, go any further in interpreting the model results directly, as all contrasts made were in relation to the baseline level of the predictors. For example, the odds ratio in line 2 of Table 2 suggests that learners were 519.32 times more likely to correctly recall the meaning of the target vocabulary at Time 2, compared to Time 1 (the baseline level of Time), but this was only when they were from the CFoF group (the baseline level of Group) and when the two continuous predictors GEVT and Listening were set at baseline (i.e., equal to zero, hence centered at mean). Therefore, in order to give a clearer picture of the three-way interactions, we decided to break them down by running multiple pairwise comparisons using the emmeans package in R (Lenth, 2019). For each comparison, the p value was adjusted using Tukey to avoid Type I errors.

The first step was to interpret the Time $\times$ GEVT $\times$ Listening interactions, which addresses the first research question. We calculated multiple comparisons between the three time points while setting the GEVT scores constant at -2 (two standard deviations below the mean), 0 (averaged at the mean), and 2 (two standard deviations above the mean), and the Listening scores constant at -2, 0, and 2. The effect plot for these interactions is given in Figure 4. We decided to set both GEVT and Listening between -2 and 2 because this range would be large enough to cover most of the observations in our dataset (i.e., from learners with very low PVK/Listening to learners with very high PVK/Listening) and would provide a clear picture of how learners with different levels of GEVT or Listening progressed in their vocabulary learning. Table 3 shows the full results

Table 2 Results for the final converged mixed effects model

	Contrast	95% CI for odds ratio			Z	p
		LL	Odds ratio	UL		
1	Intercept	0.01	**0.02**	0.04	−10.80	<.001
2	Time$_{Time2 \to Time1}$	246.12	**519.32**	1,095.79	16.41	<.001
3	Time$_{Time3 \to Time1}$	25.76	**51.12**	101.45	11.25	<.001
4	Group$_{CS \to CFoF}$	2.10	**3.25**	5.05	5.26	<.001
5	Group$_{L2 \to CFoF}$	1.01	**1.67**	2.76	1.99	.047
6	Group$_{NE \to CFoF}$	0.23	**0.43**	0.81	−2.62	.009
7	Listening	0.70	**0.97**	1.35	−0.18	.86
8	GEVT	0.93	**1.25**	1.68	1.47	.14
9	Time$_{Time2 \to Time1}$ × Group$_{CS \to CFoF}$	0.10	**0.14**	0.19	−11.78	<.001
10	Time$_{Time3 \to Time1}$ × Group$_{CS \to CFoF}$	0.06	**0.08**	0.12	−15.87	<.001
11	Time$_{Time2 \to Time1}$ × Group$_{L2 \to CFoF}$	0.03	**0.05**	0.06	−17.71	<.001
12	Time$_{Time3 \to Time1}$ × Group$_{L2 \to CFoF}$	0.09	**0.12**	0.16	−12.99	<.001
13	Time$_{Time2 \to Time1}$ × Group$_{NE \to CFoF}$	0.01	**0.02**	0.03	−19.70	<.001
14	Time$_{Time3 \to Time1}$ × Group$_{NE \to CFoF}$	0.003	**0.005**	0.009	−20.01	<.001
15	Time$_{Time2 \to Time1}$ × Listening	1.25	**1.64**	2.16	3.54	<.001
16	Time$_{Time3 \to Time1}$ × Listening	0.74	**0.94**	1.20	−0.52	.61
17	Group$_{CS \to CFoF}$ × Listening	0.62	**0.96**	1.49	−0.17	.86
18	Group$_{L2 \to CFoF}$ × Listening	0.64	**0.97**	1.47	−0.15	.88
19	Group$_{NE \to CFoF}$ × Listening	0.70	**1.07**	1.63	0.30	.76

(Continued)

Table 2 Continued

	Contrast	95% CI for odds ratio			Z	p
		LL	Odds ratio	UL		
20	Time$_{Time2→Time1}$ × GEVT	0.73	**0.92**	1.17	−0.66	.51
21	Time$_{Time3→Time1}$ × GEVT	0.82	**1.02**	1.28	0.18	.86
22	Group$_{CS→CFoF}$ × GEVT	0.80	**1.23**	1.91	0.94	.35
23	Group$_{L2→CFoF}$ × GEVT	0.87	**1.29**	1.92	1.27	.20
24	Group$_{NE→CFoF}$ × GEVT	0.57	**0.87**	1.34	−0.62	.54
25	Listening × GEVT	0.89	**1.05**	1.23	0.57	.57
26	Time$_{Time2→Time1}$ × Group$_{CS→CFoF}$ × Listening	0.65	**0.93**	1.31	−0.43	.67
27	Time$_{Time3→Time1}$ × Group$_{CS→CFoF}$ × Listening	1.09	**1.49**	2.05	2.49	.013
28	Time$_{Time2→Time1}$ × Group$_{L2→CFoF}$ × Listening	0.54	**0.75**	1.05	−1.67	.094
29	Time$_{Time3→Time1}$ × Group$_{L2→CFoF}$ × Listening	0.94	**1.28**	1.74	1.59	.11
30	Time$_{Time2→Time1}$ × Group$_{NE→CFoF}$ × Listening	0.43	**0.62**	0.87	−2.72	.007
31	Time$_{Time3→Time1}$ × Group$_{NE→CFoF}$ × Listening	0.79	**1.18**	1.78	0.81	.42
32	Time$_{Time2→Time1}$ × Group$_{CS→CFoF}$ × GEVT	0.49	**0.68**	0.95	−2.26	.024
33	Time$_{Time3→Time1}$ × Group$_{CS→CFoF}$ × GEVT	0.45	**0.62**	0.86	−2.89	.004
34	Time$_{Time2→Time1}$ × Group$_{L2→CFoF}$ × GEVT	0.74	**1.00**	1.36	0.03	.98
35	Time$_{Time3→Time1}$ × Group$_{L2→CFoF}$ × GEVT	0.63	**0.85**	1.15	−1.02	.31
36	Time$_{Time2→Time1}$ × Group$_{NE→CFoF}$ × GEVT	0.64	**0.92**	1.31	−0.46	.65
37	Time$_{Time3→Time1}$ × Group$_{NE→CFoF}$ × GEVT	0.49	**0.80**	1.29	−0.91	.36
38	Time$_{Time2→Time1}$ × Listening × GEVT	0.77	**0.87**	0.99	−2.13	.034
39	Time$_{Time3→Time1}$ × Listening × GEVT	0.80	**0.90**	1.02	−1.67	.10

Note. CI = confidence interval; LL = lower limit; UL = upper limit; CS = codeswitching; CFoF = contrastive focus-on-form; L2 = second language; NE = no explanations; GEVT = general vocabulary knowledge test.
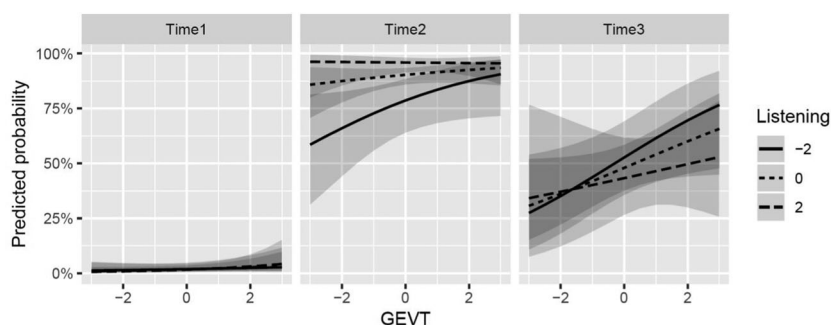
Figure 4 Effect plot for Time × GEVT (General Vocabulary Knowledge Test).

for all pairwise comparisons, but our discussion will focus on the Time2–Time1 and Time3–Time1 contrasts, because Time 1 was the baseline level of Time (pretest), and these two contrasts are of primary interest in indicating short-term learning and long-term learning, respectively.

Interpreting the results, we start with the Time2–Time1 contrasts, which indicated short-term learning. First, learners at all levels of GEVT and listening made significant short-term gains in scores on the vocabulary test. The odds ratios for these contrasts showed that the largest short-term gains were made by learners with the highest listening proficiency but with the lowest level of GEVT. In addition, learners with both the lowest GEVT scores and the lowest listening proficiency made the smallest short-term gains. Regarding long-term retention (Time3–Time1 contrasts), results were somewhat similar to what we discovered for the Time2–Time1 contrasts. The greatest long-term gains were again observed for learners with the highest listening proficiency but with the lowest GEVT level. The smallest gains, which were not statistically significant, were made by learners with the highest GEVT scores and the highest Listening proficiency.

Further rather complex but consistent patterns across the four groups emerged for short- and long-term vocabulary learning in relation to GEVT and listening levels (both Time2–Time1 and Time3–Time1 contrasts). First, the lowest level listeners (z score at -2) benefited more as their GEVT levels increased. With average to high level listeners (z scores at 0 and 2), however, their vocabulary gains decreased with an increase in their GEVT scores. Second, both middle and lower level GEVT learners (z scores at 0 and -2) made more gains on the vocabulary tests as their listening proficiency increased. This was also true for the short-term learning of the higher

Table 3 Pairwise comparisons between Time by GEVT (General Vocabulary Knowledge Test; at −2, 0, 2) and Listening (at −2, 0, 2)

GEVT	Listening	Contrast	95% CI for odds ratio			SE	z	p
			LL	Odds ratio	UL			
−2	−2	Time2–Time1	9.07	25.66	72.58	11.38	7.31	<.001
		Time3–Time1	1.16	3.21	8.89	1.40	2.68	.020
		Time3–Time2	0.07	0.12	0.24	0.03	−7.62	<.001
	0	Time2–Time1	31.46	77.92	193.00	30.15	11.26	<.001
		Time3–Time1	2.72	6.47	15.41	2.40	5.04	<.001
		Time3–Time2	0.05	0.08	0.13	0.02	−13.02	<.001
	2	Time2–Time1	65.36	236.67	856.98	129.94	9.96	<.001
		Time3–Time1	3.68	13.06	46.35	7.06	4.76	<.001
		Time3–Time2	0.02	0.06	0.14	0.02	−7.37	<.001
0	−2	Time2–Time1	12.32	29.82	72.20	11.25	9.00	<.001
		Time3–Time1	1.43	3.33	7.73	1.20	3.34	.002
		Time3–Time2	0.07	0.11	0.17	0.02	−12.33	<.001
	0	Time2–Time1	22.67	52.72	122.60	18.98	11.01	<.001
		Time3–Time1	2.01	4.41	9.68	1.48	4.42	<.001
		Time3–Time2	0.06	0.08	0.11	0.01	−18.53	<.001
	2	Time2–Time1	37.98	93.20	228.71	35.70	11.84	<.001
		Time3–Time1	2.49	5.84	13.69	2.12	4.86	<.001
		Time3–Time2	0.04	0.06	0.10	0.01	−15.12	<.001
2	−2	Time2–Time1	12.24	34.66	98.09	15.38	7.99	<.001
		Time3–Time1	1.23	3.45	9.69	1.52	2.81	.014
		Time3–Time2	0.05	0.10	0.20	0.03	−7.81	<.001
	0	Time2–Time1	14.66	35.66	86.75	13.53	9.42	<.001
		Time3–Time1	1.27	3.00	7.10	1.10	2.99	.008
		Time3–Time2	0.05	0.08	0.13	0.02	−12.67	<.001
	2	Time2–Time1	12.52	36.70	107.58	16.84	7.85	<.001
		Time3–Time1	0.93	2.61	7.35	1.15	2.17	.076
		Time3–Time2	0.03	0.07	0.15	0.02	−8.66	<.001

Note. CI = confidence interval; LL = lower limit; UL = upper limit.

GEVT learners (*z* score at 2). The long-term learning of these learners, however, benefited more with every decrease of one *SD* in their listening proficiency.

Turning to our second research question, comparing the effect of PVK and listening proficiency respectively on vocabulary learning for each of the four groups, we first obtained and then plotted Time × Group × GEVT interactions
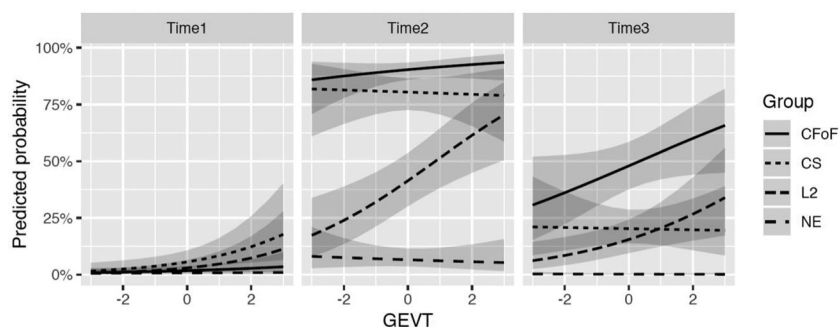
Figure 5 Effect plot for Time × Group × GEVT (General Vocabulary Knowledge Test). CFoF = contrastive focus-on-form; CS = codeswitching; L2 = second language; NE = no explanations.

(Figure 5). Then, while holding the Listening score constant at 0, we ran multiple comparisons between the three test time points by Group at three levels of GEVT: −2, 0, and 2. Table 4 shows the full results, but our discussion will again focus on the contrasts between Time2 and Time1 and between Time3 and Time1.

Results in Table 4 indicate that for short-term learning (Time2–Time1 contrasts) at all GEVT levels, all groups made significant gains, with the greatest gains observed in the CFoF group and the smallest gains in the NE group, as indicated by the odds ratios. Comparing odds ratios for the contrasts between different GEVT levels also showed that with every increase of one *SD* in learners' GEVT level, the gains on the vocabulary tests became smaller. This was the case for all groups. For example, CFoF group learners with lower GEVT scores (*z* score at −2) were 608.89 times more likely to successfully recall the meaning of the target vocabulary at Time2 compared to Time1. In contrast, their counterparts with average or higher GEVT levels (*z* score at 0 or 2, respectively) were only 519.32 and 442.93 times, respectively, more likely to do so at Time 2 compared to Time 1. In addition, although the CS approach showed a relatively bigger advantage over the L2 approach for lower level GEVT learners (odds ratio: 177.41 vs. 27.35), with every increase of one *SD* in learners' GEVT level, this advantage seemed to get smaller. Indeed, when learners' GEVT score reached 2, the odds ratio for the CS group contrast was only slightly higher than that for the L2 group (27.72 vs. 20.25).

Regarding differences in long-term vocabulary retention (Time3–Time1 contrasts) between groups, again the CFoF group made the largest gains at

Table 4 Pairwise comparisons between Time by Group at three levels of GEVT (General Vocabulary Knowledge Test; −2, 0, and 2)

GEVT	Contrast	Group	LL	Odds ratio	UL	SE	z	p
				95% CI for odds ratio				
−2	Time2–Time1	CFoF	207.94	608.89	1,782.99	279.13	13.99	<.001
	Time3–Time1	CFoF	18.04	49.05	133.40	20.94	9.12	<.001
	Time3–Time2	CFoF	0.05	0.08	0.14	0.02	−10.43	<.001
	Time2–Time1	CS	61.04	177.41	515.69	80.77	11.37	<.001
	Time3–Time1	CS	3.93	10.64	28.79	4.52	5.57	<.001
	Time3–Time2	CS	0.03	0.06	0.11	0.02	−10.23	<.001
	Time2–Time1	L2	9.96	27.35	75.11	11.79	7.68	<.001
	Time3–Time1	L2	3.01	7.99	21.20	3.33	4.99	<.001
	Time3–Time2	L2	0.17	0.29	0.51	0.07	−5.17	<.001
	Time2–Time1	NE	4.09	12.48	38.06	5.94	5.31	<.001
	Time3–Time1	NE	0.11	0.42	1.58	0.24	−1.53	.28
	Time3–Time2	NE	0.01	0.03	0.10	0.02	−7.58	<.001
0	Time2–Time1	CFoF	212.64	519.32	1,268.29	197.85	16.41	<.001
	Time3–Time1	CFoF	22.52	51.12	116.02	17.88	11.25	<.001
	Time3–Time2	CFoF	0.07	0.10	0.14	0.01	−15.34	<.001
	Time2–Time1	CS	29.58	70.12	166.22	25.82	11.54	<.001
	Time3–Time1	CS	1.95	4.33	9.62	1.48	4.30	<.001
	Time3–Time2	CS	0.04	0.06	0.09	0.01	−18.88	<.001
	Time2–Time1	L2	9.86	23.53	56.14	8.73	8.51	<.001
	Time3–Time1	L2	2.70	6.08	13.69	2.11	5.21	<.001
	Time3–Time2	L2	0.18	0.26	0.37	0.04	−9.15	<.001
	Time2–Time1	NE	3.65	9.01	22.26	3.48	5.70	<.001
	Time3–Time1	NE	0.11	0.28	0.72	0.11	−3.18	.004
	Time3–Time2	NE	0.02	0.03	0.06	0.01	−14.02	<.001
2	Time2–Time1	CFoF	156.58	442.93	1,252.96	196.52	13.73	<.001
	Time3–Time1	CFoF	20.53	53.27	138.22	21.67	9.77	<.001
	Time3–Time2	CFoF	0.07	0.12	0.21	0.03	−9.05	<.001
	Time2–Time1	CS	10.26	27.72	74.91	11.76	7.83	<.001
	Time3–Time1	CS	0.70	1.76	4.47	0.70	1.43	.33
	Time3–Time2	CS	0.04	0.06	0.11	0.02	−11.21	<.001
	Time2–Time1	L2	7.73	20.25	53.01	8.31	7.32	<.001
	Time3–Time1	L2	1.85	4.63	11.58	1.81	3.92	<.001
	Time3–Time2	L2	0.14	0.23	0.38	0.05	−6.81	<.001

(*Continued*)

Table 4 Continued

GEVT	Contrast	Group	95% CI for odds ratio			SE	z	p
			LL	Odds ratio	UL			
	Time2–Time1	NE	1.97	6.51	21.47	3.31	3.68	.001
	Time3–Time1	NE	0.04	0.19	0.85	0.12	−2.60	.025
	Time3–Time2	NE	0.01	0.03	0.10	0.02	−6.59	<.001

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit; CFoF = contrastive focus-on-form; CS = codeswitching; L2 = second language; NE = no explanations.

all levels of GEVT, with the largest odds ratios observed. Unlike what was found for short-term learning, the NE group at the lower GEVT level made no significant progress, with significant decreases in vocabulary scores at the average and higher GEVT levels. In addition, comparing odds ratios for the contrasts between different GEVT levels showed that the gains for learners in the CFoF group became larger with every increase of one *SD* in learners' GEVT scores (49.05 vs. 51.12 vs. 53.27). The gains for the CS and L2 groups, however, decreased with every increase of learners' GEVT levels. Finally, a further comparison of odds ratios between the CS and L2 groups indicated that when learners' GEVT was at the lower level (−2), they benefited more from the CS approach than the L2 approach. Learners who were at the average (0) and higher (2) GEVT levels were helped more by the L2 approach than by the CS approach.

For listening, we followed a similar procedure to that for the Time × Group × GEVT interactions. We first plotted the Time × Group × Listening interaction (Figure 6) and then ran multiple comparisons between the three test time points by Group at three levels of Listening: −2, 0, 2. Table 5 presents the full results for all pairwise comparisons, but our discussion will again focus on contrasts between Time2 and Time1 and between Time3 and Time1.

The Time2–Time1 contrasts listed in Table 5 showed that learners across the three Listening levels from all four groups made significant short-term vocabulary gains. The odds ratios for these contrasts first indicated that, at all three Listening levels, the greatest vocabulary gains were made by the CFoF group, followed by the CS, L2, and finally the NE groups. Second, vocabulary gains across the four groups became larger with every increase of one *SD* in learners' Listening proficiency. Regarding long-term vocabulary retention (Time3–Time1 contrasts), it was only in the CFoF and L2 groups that learners at all three Listening proficiency levels made significant improvement. Whereas

Table 5 Pairwise comparisons between Time by Group at three levels of Listening (−2, 0, and 2)

Listening	Contrast	Group	95% CI for odds ratio			SE	z	p
			LL	Odds ratio	UL			
−2	Time2–Time1	CFoF	65.65	192.62	565.14	88.46	11.45	<.001
	Time3–Time1	CFoF	21.72	58.08	155.30	24.37	9.68	<.001
	Time3–Time2	CFoF	0.17	0.30	0.55	0.08	−4.70	<.001
	Time2–Time1	CS	11.22	30.28	81.76	12.83	8.05	<.001
	Time3–Time1	CS	0.86	2.21	5.65	0.88	1.97	.12
	Time3–Time2	CS	0.04	0.07	0.13	0.02	−10.82	<.001
	Time2–Time1	L2	5.59	15.35	42.13	6.61	6.34	<.001
	Time3–Time1	L2	1.59	4.20	11.11	1.74	3.46	.002
	Time3–Time2	L2	0.15	0.27	0.49	0.07	−5.30	<.001
	Time2–Time1	NE	3.14	8.83	24.85	3.90	4.94	<.001
	Time3–Time1	NE	0.07	0.23	0.77	0.12	−2.84	.013
	Time3–Time2	NE	0.01	0.03	0.07	0.01	−9.09	<.001
0	Time2–Time1	CFoF	212.64	519.32	1,268.29	197.85	16.41	<.001
	Time3–Time1	CFoF	22.52	51.12	116.02	17.88	11.25	<.001
	Time3–Time2	CFoF	0.07	0.10	0.14	0.01	−15.34	<.001
	Time2–Time1	CS	29.58	70.12	166.22	25.82	11.54	<.001
	Time3–Time1	CS	1.95	4.33	9.62	1.48	4.30	<.001
	Time3–Time2	CS	0.04	0.06	0.09	0.01	−18.88	<.001
	Time2–Time1	L2	9.86	23.53	56.14	8.73	8.51	<.001
	Time3–Time1	L2	2.70	6.08	13.69	2.11	5.21	<.001
	Time3–Time2	L2	0.18	0.26	0.37	0.04	−9.15	<.001
	Time2–Time1	NE	3.65	9.01	22.26	3.48	5.70	<.001
	Time3–Time1	NE	0.11	0.28	0.72	0.11	−3.18	.004
	Time3–Time2	NE	0.02	0.03	0.06	0.01	−14.02	<.001
2	Time2–Time1	CFoF	447.98	1,400.13	4,376.04	680.79	14.90	<.001
	Time3–Time1	CFoF	16.16	44.99	125.26	19.66	8.71	<.001
	Time3–Time2	CFoF	0.02	0.03	0.06	0.01	−11.85	<.001
	Time2–Time1	CS	57.06	162.38	462.10	72.46	11.41	<.001
	Time3–Time1	CS	3.21	8.50	22.49	3.53	5.15	<.001
	Time3–Time2	CS	0.03	0.05	0.10	0.01	−11.02	<.001
	Time2–Time1	L2	13.69	36.08	95.10	14.92	8.67	<.001
	Time3–Time1	L2	3.50	8.80	22.15	3.47	5.52	<.001

(*Continued*)

Table 5 Continued

Listening	Contrast	Group	95% CI for odds ratio			SE	z	p
			LL	Odds ratio	UL			
	Time3–Time2	L2	0.15	0.24	0.40	0.05	−6.57	<.001
	Time2–Time1	NE	3.21	9.20	26.37	4.13	4.94	<.001
	Time3–Time1	NE	0.10	0.35	1.18	0.18	−2.03	.11
	Time3–Time2	NE	0.01	0.04	0.10	0.02	−8.18	<.001

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit; CFoF = contrastive focus-on-form; CS = codeswitching; L2 = second language; NE = no explanations.
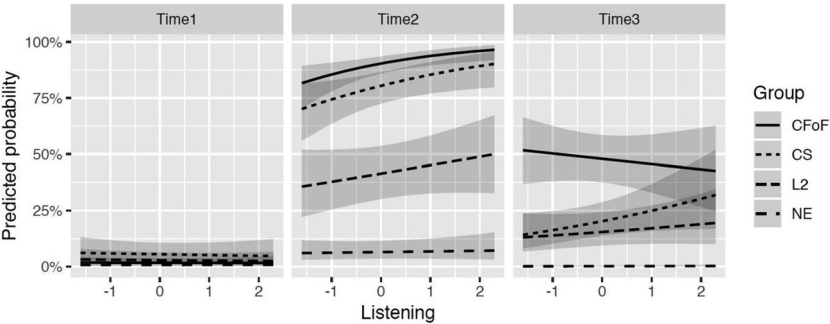
Figure 6 Effect plot for Time × Group × Listening. CFoF = contrastive focus-on-form; CS = codeswitching; L2 = second language; NE = no explanations.

learners at both average and higher Listening proficiency levels in the CS group made significant gains, lower level listeners did not. Additionally, higher level listeners from the NE group made no significant gains, and average and less proficient listeners from the NE group showed a significant decrease in vocabulary scores between Time 1 and Time 3. Furthermore, although the lower and average level listeners in the L2 group showed larger gains than their counterparts in the CS group, higher level listeners showed similar progress across both groups, with very similar odds ratios observed (8.80 vs. 8.50).

Discussion

We summarize our findings as follows.

For our first research question, examining the impact of learners' PVK and listening proficiency on vocabulary learning through listening for all conditions together, we found that learners' listening proficiency played a more important

Table 6 Variation in impact of different vocabulary explanations according to level of preexisting vocabulary knowledge (PVK) and listening proficiency, organized by proficiency level

Proficiency	Short-term gains	Long-term gains
Higher PVK	CFoF > CS > L2 > NE	CFoF > L2 > CS > NE
Lower PVK	CFoF > CS > L2 > NE	CFoF > CS > L2 > NE
Higher listening	CFoF > CS > L2 > NE	CFoF > CS = L2 > NE
Lower listening	CFoF > CS > L2 > NE	CFoF > L2 > CS > NE

Note. CFoF = contrastive focus-on-form; CS = codeswitching; L2 = second language; NE = no explanations.

Table 7 Variation in impact of different vocabulary explanations according to level of preexisting vocabulary knowledge (PVK) and listening proficiency, organized by vocabulary explanation group

Group	Short-term gains	Long-term gains
CFoF	Lower PVK > Higher PVK Higher listening > Lower listening	Higher PVK > Lower PVK Lower listening > Higher listening
CS, L2, and NE	Lower PVK > Higher PVK Higher listening > Lower listening	Lower PVK > Higher PVK Higher listening > Lower listening

Note. CFoF = contrastive focus-on-form; CS = codeswitching; L2 = second language; NE = no explanations.

role than PVK levels did, with the largest short-term and long-term vocabulary gains made by learners with the highest level listening proficiency but with the lowest level of PVK. In addition, whereas the lower level listeners made larger vocabulary gains with every increase of one *SD* in their PVK level, vocabulary learning gains for the average and higher level listeners decreased with every increase in their PVK level.

Our second research question asked whether the impact of post-listening vocabulary explanations in the L2, CS, CFoF, and NE groups varied according to learners' PVK or listening proficiency (summarized in Tables 6 and 7).

Turning first to PVK, our findings suggested that the CFoF approach was the most beneficial for short- and long-term vocabulary learning for learners across the PVK levels. Although lower PVK learners in the CFoF group made greater short-term gains than higher PVK learners did, the long-term learning gains from the CFoF group were greater for higher PVK learners. In other

words, there were benefits for both PVK levels. These findings contrast with those for the NE group, which was the least beneficial teaching approach across all PVK levels for vocabulary learning.[2] Furthermore, within the CS, L2, and NE groups, the higher the learners' PVK level was, the smaller the short-term vocabulary gains they made. The same was true for long-term learning within the CS and L2 groups; for the NE group, no significant gains were made by learners at any PVK level. Finally, differences emerged between the L2 and CS groups. Whereas higher PVK learners made greater gains in the L2 approach, lower level PVK learners benefited more from the CS approach.

Regarding whether the impact of post-listening vocabulary explanations in the L2, CS, CFoF, and NE groups varied according to learners' listening proficiency, findings were in some ways similar to what was found for PVK. First, CFoF again emerged as the most beneficial approach for both short-term and long-term learning for learners across the listening proficiency levels. Also similar to what was found for PVK, the NE teaching approach was the least beneficial across listening proficiency levels.

Second, and different from what was found for PVK, for short-term learning within the CS, L2, and NE groups, the higher the learners' listening level was, the larger the gains they made in vocabulary learning. The same was true for long-term learning within the CS and L2 groups, with no significant long-term learning gains for the NE group at any of the three listening proficiency levels.

Third, differences again emerged between the L2 and CS groups. Learners' listening proficiency seemed to be a stronger predictor for vocabulary learning in the CS group than in the L2 group. Learners at all listening proficiency levels benefited more from the CS approach than from the L2 approach for short-term learning, but the effect was most marked for the higher proficiency listeners. By contrast, for long-term learning, both the CS and L2 approaches led to similar gains for higher listening proficiency learners, whereas the L2 approach led to greater long-term vocabulary gains for lower and average level listeners than the CS.

We interpret these complex findings as follows. First, the more important role of listening proficiency compared with PVK (as evidenced by the Time $\times$ GEVT $\times$ Listening interaction) suggests that these two variables each play a different role in vocabulary learning through aural input and that the relationship between them is not necessarily a straightforward one (Smith, 2019). In other words, learners' vocabulary size may not be a wholly reliable indication of their ability to understand spoken input. Furthermore, even though we assessed PVK through an aural test, that test might not have fully captured learners' ability to recognize vocabulary in connected speech (van Zeeland, 2017).

Second, the fact that learners with the highest listening proficiency but the lowest PVK made the greatest overall vocabulary gains suggests that the ability to understand spoken input (as measured by our comprehension tests) despite lower PVK brings benefits for learning. Such learners resemble those in the studies by van Zeeland and Schmitt (2013b) and Bonk (2000), who achieved higher levels of listening comprehension than might have been expected from their vocabulary size. Commenting that "some L2 listeners seem to cope better with unknown vocabulary than others" (p. 473), van Zeeland and Schmitt also acknowledge the possibility of those learners having "better metacognitive control of the comprehension process using effective combinations of different cognitive processes such as inferencing, elaborating, monitoring, evaluating, and predicting" (p. 474). In turn, that approach, which seems to involve more active and strategic engagement with the listening input, may lead to greater vocabulary learning than was the case for those in our study who had higher levels of both PVK and listening proficiency. Learners with lower levels of listening proficiency, by contrast, made greater gains as their PVK level increased. That suggests that where listening proficiency was less developed, more PVK was needed to enable vocabulary learning from aural input. In such cases, PVK perhaps facilitated inferencing and the use of other strategies that helped learners to work out the meaning of any unknown words in the input, but these processes functioned less effectively when PVK was too low (see Graham, Santos, & Vanderplank, 2010, for similar arguments). At higher levels of listening proficiency, however, higher PVK may lead to smaller vocabulary gains, arguably because learners are having to work less hard to understand the input and therefore process it less deeply.

Looking at the impact of PVK in relation to each of the different types of vocabulary explanations, CFoF emerged as the most balanced approach, that is, with the greatest learning gains regardless of learners' PVK level. The crosslinguistic information provided about vocabulary items may have shifted the attentional direction (Ellis, 1999) of learners in the CFoF group away from the general meaning of the teacher's explanations to the target words themselves, made more salient and hence encouraging greater noticing (Schmidt, 1990). It may also have encouraged deeper processing and greater evaluation of the target items in so far as their uses in the L1 and L2 were compared and contrasted. The approach thus seemed to prompt greater involvement load as described by Hulstijn and Laufer (2001) and thence better and more durable learning. Lower PVK learners in the CFoF group, however, may have had more limited understanding of the crosslinguistic information provided, meaning that the gains made were short-term rather than long-term.

The benefits of the CFoF approach for higher PVK learners stands in contrast to how they fared in the L2, CS, and NE groups, where higher PVK levels were associated with smaller learning gains, perhaps because those conditions did not provide the additional linguistic information that seemed to be helpful for higher PVK learners in the CFoF condition. The CS condition, furthermore, was less helpful for higher PVK learners than the L2 approach was for long-term learning. Learners with a lower level of PVK, however, benefited more from the CS approach than from the L2 approach. This last finding is in line with Lee and Levine (2020), who found that intermediate learners in their study benefited significantly more from CS. Both findings would be supported by Kroll and Stewart's (1994) revised hierarchical model of bilingual memory. According to that model, although L2 words need to be attached to their L1 translation before access to the underlying mental concepts occurs (i.e., at lower proficiency levels), direct conceptual links between L2 words and conceptual representations can also be created when higher proficiency levels are reached. Learners with a larger vocabulary size in the L2 group (and also those in the CFoF group) may have gone beyond the early stage of language learning and been able to establish direct conceptual links between the target L2 words and their concepts, therefore having better understanding of the L2 explanations and the crosslinguistic explanations and hence experiencing significantly larger vocabulary gains compared with learners who had a smaller vocabulary size.

Turning to listening proficiency, the CFoF approach was again the most beneficial for vocabulary learning, regardless of proficiency level. The long-term learning benefits for the lower listening proficiency learners in the CFoF condition may have arisen because they gained from the crosslinguistic explanations the kind of information that they were less able to extract for themselves from the listening passage alone. That may have been particularly true for those learners who had lower listening proficiency but higher PVK; as noted above, higher PVK was linked with better long-term learning in the CFoF condition. The short-term learning gains for higher listening proficiency learners in the CFoF group may similarly relate back to the finding that learners with higher listening proficiency and lower PVK made the greatest gains overall, but that lower PVK was associated with short-term gains in the CFoF group.

In the other conditions, however, in contrast to what was found for PVK, the higher learners' listening proficiency was before the intervention, the greater the short-term vocabulary gains they made. For the CS group, where learners only had access to the teacher's L1 translation of vocabulary items, higher listening proficiency may have allowed them to supplement such basic information with

information gained from the listening passage regarding the items' use. That may also have been the case for higher listening proficiency learners in the NE group, who had NE of vocabulary items at all. For the L2 group, higher listening proficiency probably not only facilitated greater understanding of the teacher's explanations, but also perhaps allowed them to extract information from the passage itself and the example sentence provided by the teacher that presented target items in context.

These findings therefore, although confirming the benefits of the CFoF approach for vocabulary learning reported in Zhang and Graham (2019), constitute an important extension to that study, by showing how forms of instruction for aural input interact with learners' PVK and listening proficiency levels.

Limitations and Future Directions

Our study was set in a high school in China, where it is not unusual for teachers to provide vocabulary explanations alongside meaning-focused comprehension activities, as also seems to be true in other contexts with learners of a similar age (Pujadas & Muñoz, 2019). We acknowledge, however, that our findings may not extend to contexts where listening to aural input is less closely directed by a teacher. Similarly, in our study each group only experienced one form of treatment and learners had to remain in intact classes, as is often the case in school-based investigations, thus not permitting random assignment to different treatment groups. This poses a limitation. The same is true of the use of a different test format at pretest compared with posttest and delayed posttest (see the section above Vocabulary Posttests and Delayed Posttests). We addressed these limitations, however, first by using generalized linear mixed effects models to analyze the data, which controlled random effects due to repeated testing and individual differences. Second, allowing learners' PVK and listening proficiency, assessed at baseline, to interact with the other two predictors (test time points and treatment conditions) in the analysis further addressed the possible limitations arising from the lack of random assignment of students to different treatment groups. Furthermore, our findings, using PVK and listening as separate, continuous variables indexing general language proficiency, provide a clearer picture of how both variables interact with vocabulary learning through listening than previous studies using categorical variables and combined, general proficiency measures (e.g., Hennebry et al., 2013; Lee & Levine, 2020; Tian & Macaro, 2012; Vidal, 2011).

We can therefore conclude with some confidence that a CFoF approach is the most helpful type of post-listening teacher explanation for learners regardless of their PVK and listening proficiency. Comparing just the CS and L2 groups,

the CS approach was more helpful for lower PVK learners and more proficient listeners than the L2 approach was. Learners with a higher level of PVK and those with a lower level of listening proficiency, however, benefited more from the L2 approach than from the CS approach. Our potentially most interesting finding, however, relates to the importance of listening proficiency as a factor influencing vocabulary learning through aural input. It would be useful for future research to explore exactly how important it is. Future research might also consider whether the impact of listening proficiency, PVK, and teaching approach is similar for both collocations and single words.

In addition, the fact that vocabulary gains were greatest in the long-term for learners with higher listening proficiency and lower PVK is worthy of further exploration, to establish why and how such students are able to learn particularly well from aural input. These findings provide important evidence of the complex interplay of vocabulary knowledge, listening proficiency, and instructional conditions, with implications for theories of vocabulary learning. They underline the potential relevance of the involvement load hypothesis (Hulstijn & Laufer, 2001) for learning through aural input, an issue largely overlooked by previous research. That is, our findings point to the importance of conditions, such as explicitly contrasting the L2 forms with the equivalent L1 forms, that seem to facilitate deeper processing and hence better learning. For vocabulary learning during listening, this may be aural input and learning activities that require some degree of effortful or strategic engagement without being so far beyond learners' level of vocabulary knowledge and listening proficiency that such engagement is inhibited.

Conclusion

In conclusion, these findings, which are of relevance beyond the Chinese learners of English we studied, add to our understanding of vocabulary learning through listening, of the potential value of CFoF, and of the role of listening proficiency and prior vocabulary knowledge within such approaches. At a pedagogical level, learners' vocabulary size and listening proficiency should be taken into consideration when planning activities to enhance vocabulary learning through listening, with due attention given to the benefits that might be gained from an approach using CFoF.

<div style="text-align: right">Final revised version accepted 3 February 2020</div>

Notes

1 The vocabulary posttests and delayed posttests were scored either right or wrong. We acknowledge the fact that these tests could have been scored ordinally (e.g., by giving half marks), as suggested by a reviewer. However, scoring ordinally would have required us to use ordinal logistic regression. We already had a rather complex fixed effects structure with three-way interactions between predictor variables in the generalized linear mixed effects models; adding an additional level to the outcome variable would, we believe, have overcomplicated the statistical analyses.
2 As commented by a reviewer, lower levels of vocabulary learning for the NE group are perhaps unsurprising; that this was true regardless of PVK is still noteworthy, however.

Open Research Badges

This article has earned an Open Materials badge for making publicly available the components of the research methods needed to reproduce the reported procedure. All materials that the authors have used and have the right to share are available at https://www.iris-database.org/. All proprietary materials have been precisely identified in the manuscript.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bonk, W. J. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening*, *14*, 14–31. https://doi.org/10.1080/10904018.2000.10499033

Breeze, R., & Miller, P. (2012). Predictive validity of the IELTS listening test as an indicator of student coping ability in Spain. In L. Taylor & C. J. Weir (Eds.), *IELTS collected papers 2* (pp. 487–518). Cambridge, UK: Cambridge University Press.

British Council. (n.d.). *IELTS*. Retrieved January 22, 2019, from Retrieved from https://takeielts.britishcouncil.org/find-out-about-results/understand-your-ielts-scores

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading a Foreign Language*, *20*, 136–163.

Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing*, *35*, 3–25. https://doi.org/10.1177/0265532216676851

Cobb, T. (n.d.). VocabProfilers. October 20, 2019. Retrieved from http://www.lextutor.ca/vp/

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge, UK: Press Syndicate of the University of Cambridge.

Ellis, R. (1999). Factors in the incidental acquisition of second language vocabulary from oral input. In R. Ellis (Ed.), *Learning a second language through interaction* (pp. 35–62). Amsterdam, Netherlands: John Benjamins.

Graham, S., Santos, D., & Vanderplank, R. (2010). Strategy clusters and sources of knowledge in French L2 listening comprehension. *International Journal of Innovation in Language Learning and Teaching*, *4*, 1–20. https://doi.org/10.1080/17501220802385866

Hennebry, M., Rogers, V., Macaro, E., & Murphy, V. (2013). Direct teaching of vocabulary after listening: Is it worth the effort and what method is best? *Language Learning Journal*, *45*, 282–300. https://doi.org/10.1080/09571736.2013.849751

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge, UK: Cambridge University Press.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, *51*, 539–558. https://doi.org/10.1111/0023-8333.00164

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, *24*, 617–637. https://doi.org/10.1017/S0272263102004047

Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal*, *88*, 416–432. https://doi.org/10.1111/j.0026-7902.2004.00238.x

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149–174. https://doi.org/10.1006/jmla.1994.1008

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Laufer, B. (2005). Focus on form in second language vocabulary learning. *Eurosla Yearbook*, *5*, 223–250. https://doi.org/10.1075/eurosla.5.11lau

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, *29*, 694–716. https://doi.org/10.1093/applin/amn018

Lee, J. H., & Levine, G. S. (2020). The effects of instructor language choice on second language vocabulary learning and listening comprehension. *Language Teaching Research*, *24*, 250–272. https://doi.org/10.1177/1362168818770910

Lenth, R. (2019). Emmeans: Estimated marginal means, aka least-squared means. Retrieved from https://CRAN.R-project.org/package=emmeans

Liu, D., Gong, Y., Zheng, W., Brooks, D., Allingham, N., Sjoquist, R., & Miller, S. (2007). *New senior English for China*. Beijing, China: People's Education Press.

McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, *19*, 741–760. https://doi.org/10.1177/1362168814567889

Ministry of Education, P. R. China. (2003). *Putong gaozhong yingyu kecheng biaozhun [English language curriculum guidelines for senior secondary school]*. Beijing: People's Education Press.

Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *International Review of Applied Linguistics in Language Teaching*, *56*, 1–24. https://doi.org/10.1515/iral-2017-0016

Montero Perez, M., & Desmet, P. (2012). The effect of input enhancement in L2 listening on incidental vocabulary learning: A review. *Procedia – Social and Behavioral Sciences*, *34*, 153–157. https://doi.org/10.1016/j.sbspro.2012.02.031

Nation, I. S. P. (2001). *Learning vocabulary in another language* (1st ed.). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*, 551–577. https://doi.org/10.1017/S0272263117000407

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *Language Learning Journal*, *47*, 479–496. https://doi.org/10.1080/09571736.2019.1616806

R Development Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions* (Unpublished doctoral dissertation). Victoria University, Wellington, New Zealand.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158. https://doi.org/10.1093/applin/11.2.129

Sharwood Smith, M. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research*, *7*, 118–132. https://doi.org/10.1177/026765839100700204

Silver, R. E., Hu, G., & Iino, M. (2002). *English language education in China, Japan, and Singapore*. Singapore: National Institution of Education, Nanyang Technology University.

Smith, G. (2019, March). *The relationship between L2 vocabulary knowledge and listening comprehension ability: A meta-analysis*. Paper presented at the AAAL Conference, Atlanta, Georgia.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, *36*, 139–152. https://doi.org/10.1080/09571730802389975

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, *31*, 577–607. https://doi.org/10.1017/S0272263109990039

Tauroza, S., & Allison, D. (1990). Speech rates in British English. *Applied Linguistics*, *11*, 90–105. https://doi.org/10.1093/applin/11.1.90

Tian, L. (2011). *Teacher codeswitching in a communicative EFL context: Measuring the effects on vocabulary learning*. Saarbrücken, Germany: VDM.

Tian, L., & Macaro, E. (2012). Comparing the effect of teacher codeswitching with English-only explanations on the vocabulary acquisition of Chinese university students: A Lexical Focus-on-Form study. *Language Teaching Research*, *16*, 367–391. https://doi.org/10.1177/1362168812436909

van Zeeland, H. (2017). Four studies on vocabulary knowledge in and from listening: Findings and implications for future research. *Language Teaching*, *50*, 143–150. https://doi.org/10.1017/S0261444816000318

van Zeeland, H., & Schmitt, N. (2013a). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, *41*, 609–624. https://doi.org/10.1016/j.system.2013.07.012

van Zeeland, H., & Schmitt, N. (2013b). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*, 457–479. https://doi.org/10.1093/applin/ams074

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*, 219–258. https://doi.org/10.1111/j.1467-9922.2010.00593.x

Zhang, P., & Graham, S. (2019). Vocabulary learning through listening: Comparing L2 explanations, teacher codeswitching, contrastive focus-on-form and incidental learning. *Language Teaching Research*, https://doi.org/10.1177/1362168819829022

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Example Explanations Given in the Three Treatment Groups.

Appendix S2. Listening Passage Used in the Intervention.

Appendix S3. Vocabulary Profile of the Listening Passages.

Appendix S4. Target Vocabulary Items and their Contextual Sentences.

Appendix S5. Scoring System for the Vocabulary Posttests and Delayed Posttests.

Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

How Can Teachers Best Support Second Language Vocabulary Learning Through Listening?

What This Research Was About and Why It Is Important

Having a wide vocabulary is important for all aspects of second language (L2) learning. Learners can pick up new L2 vocabulary through listening, but usually only to a limited degree. In the classroom, learning through listening can be enhanced by activities that draw learners' attention to the new vocabulary, including through explanations of new words by the teacher. However, how far learners of different L2 proficiency levels benefit from different types of explanations is unclear. This study involved high school Chinese learners of English. The researchers investigated how learners' preexisting vocabulary knowledge (PVK) and listening proficiency affected how much vocabulary they learnt from listening to passages in English, when they were followed by different types of vocabulary explanations from their teacher: NE; explanations in the L2; codeswitched explanations (using learners' first language [L1] or mother tongue); and CFoF explanations, where the teacher compared how the word functions in the L1 and L2. Findings point to the importance of listening proficiency and PVK alongside using a CFoF approach.

What the Researchers Did

- 137 learners from four classes were randomly assigned to a No Explanation group, L2 group, Codeswitching group, or CFoF group, and participated in a six-week classroom intervention.
- Before the intervention, the following measures were taken: learners' general PVK, their listening proficiency; and their knowledge of the 60 vocabulary items to be taught.
- Each group had six intervention sessions (one per week, 10 vocabulary items taught per session). In each session they listened to a passage, followed by explanations of the 10 items by their teacher.

- At the end of each session, learners completed a test on the vocabulary items taught in it. Vocabulary from all 6 sessions were then tested again two weeks later to see how much learners remembered in the longer-term.

What the Researcher Found
- Learners' listening proficiency was a more important factor influencing vocabulary learning through listening than their PVK.
- Learners who were good at listening but had lower levels of PVK learned the most words.
- Less vocabulary learning occurred when both PVK and listening proficiency were low or when both were high.
- The CFoF approach was most helpful for learners regardless of their PVK and listening proficiency, while the NE approach was the least helpful.

Things to Consider
- The CFoF approach, which draws learners' attention to key aspects of the new vocabulary and how it relates to the L1, seemed the most "balanced" approach regardless of learners' levels of PVK and listening proficiency.
- Improving learners' listening proficiency (e.g., helping them to segment words from the input or draw inferences) may help them learn vocabulary from spoken language.
- Using listening passages that offer some but not too much challenge may prompt learners to use listening strategies that might help their vocabulary learning.

Materials: Materials are freely available at https://www.iris-database.org/
How to cite this summary: Zhang, P., & Graham, S. (2020). How can teachers best support second language vocabulary learning through listening? *OASIS Summary* of Zhang & Graham (2020) in *Language Learning*. https://oasis-database.org

This summary has a CC BY-NC-SA license.