

*Vocabulary learning through listening:
comparing L2 explanations, teacher
codeswitching, contrastive focus-on form
and incidental learning*

Article

Accepted Version

Zhang, P. and Graham, S. (2020) Vocabulary learning through listening: comparing L2 explanations, teacher codeswitching, contrastive focus-on form and incidental learning. *Language Teaching Research*, 24 (6). pp. 765-784. ISSN 1477-0954 doi: <https://doi.org/10.1177/1362168819829022> Available at <https://centaur.reading.ac.uk/81387/>

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/10.1177/1362168819829022>

Publisher: Sage

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online



Vocabulary learning through listening: Comparing L2 explanations, teacher codeswitching, contrastive focus-on-form and incidental learning

Journal:	<i>Language Teaching Research</i>
Manuscript ID	LTR-18-0258.R1
Manuscript Type:	Full Research Article
Keywords:	vocabulary acquisition, listening comprehension, teacher codeswitching, second language, contrastive focus on form, incidental learning
Abstract:	<p>This study explored the teaching and learning of vocabulary through listening among 137 senior high-school EFL learners in China. It compared different types of Lexical Focus-on-Form delivered to four treatment groups: post-listening vocabulary explanations in the L2; codeswitched explanations; explanations providing additional cross-linguistic information (Contrastive Focus-on-Form, CFoF); and no explanations (NE). It also investigated the impact of the intervention on learners' listening comprehension. Learners completed aural vocabulary tests at pre-, post- and delayed post-test and listening assessments at pre- and post-test. For short and long-term vocabulary acquisition, the three groups receiving explanations significantly outperformed the NE group. Gains for the CFoF group were significantly greater than for the L2 and Codeswitching groups, for both short-term and long-term learning. For listening comprehension, only the NE group made significant improvement from the pre-test to the post-test, as well as making significantly greater pre to post-test improvement than the CFoF and the L2 groups did. The paper concludes by discussing these findings in relation to theories of vocabulary acquisition and listening comprehension, as well as their pedagogical implications.</p>

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Vocabulary learning through listening

Introduction

The development of English proficiency is an important goal for high school learners in China, as it is in many contexts. Fundamental to that goal is the acquisition of a wide vocabulary (Bogaards and Laufer, 2004). Relatively little is known, however, regarding the relative benefits to vocabulary development of different classroom practices. In addition, while current educational policy in China encourages maximum second language (L2) use by high school teachers in the classroom, it is unclear whether this is, in fact, the optimum teaching approach (Silver et al., 2002). Aiming to fill this gap, this study explores the impact of different types of vocabulary instruction, comparing teacher codeswitching, cross-linguistic explanations, and exclusive L2 use, within English as a Foreign Language (EFL) teacher-learner interactions following listening activities.

Listening is also regarded as a “foundational skill” (Gu, 2018), and vocabulary acquisition through listening is arguably of particular import given the centrality of oral input in communicative language teaching. In settings such as China, where traditionally language teaching has focused on written texts, there is a move towards a more communicative, oral-based approach (Lee, 2010), underscoring the importance of investigating how vocabulary can be best developed through oral input.

LITERATURE REVIEW

Within classroom-focused research, it is common to view L2 vocabulary learning as involving either intentional learning or incidental learning (Hulstijn, 2001). In the latter, learners learn vocabulary “as the by-product of any activity not explicitly geared to vocabulary learning”; conversely, intentional vocabulary learning refers to “any activity aiming at committing lexical information to memory” (Hulstijn, 2001: 267). Although intentional vocabulary learning can result in better short-term outcomes compared to incidental learning (Laufer, 2009), both types of learning have different advantages, and in practice the division between them is not always entirely clear-cut (Hulstijn, 2001). Furthermore, both incidental and intentional learning can be enhanced by pedagogical activities that draw learners’ attention to key information about a word, an approach termed Lexical Focus-on-Form (Laufer and Girsai, 2008). As yet, however, relatively little research has been focused on incidental learning followed by intentional vocabulary instruction (Hennebry et al., 2013).

Existing research has also tended to focus on learning through written input (Pellicer-Sánchez and Schmitt, 2010) rather than through oral input. The few studies undertaken show that vocabulary

Vocabulary learning through listening

knowledge can be acquired incidentally through oral input, but levels and rate of acquisition differ from what is found for reading. For example, Vidal (2011) compared incidental vocabulary acquisition through oral and written input among 230 university learners of English. While oral input led to the retention of a smaller number of words than was the case for reading, retention was more durable in the listening condition (except in very low proficiency learners), perhaps because oral input can pass directly into phonological memory, with accompanying “stable long-term memory representations” (Vidal, 2011: 244-245). Yet fewer words are retained than through reading, perhaps because in listening there tends to be a greater need to gain the overall gist of what is said, whereas reading “allows more attention to be focused on language form” (p. 248). The fleeting nature of oral input and learners’ difficulties with speech segmentation make it more difficult to process such input, to isolate and focus on specific lexical items. This suggests that, more so than for reading, incidental learning needs to be supplemented by intentional, explicit Lexical Focus-on Form teaching.

The role of the L1 in such teaching, however, is also underexplored, although there is emerging evidence of an advantage from the use of teacher codeswitching or CS (Tian, 2011; Tian and Macaro, 2012). Potentially better learning from teacher CS can be explained theoretically through models of vocabulary representation and development such as that of Jiang (2000) in which the L1 occupies a central position. Jiang (2000) proposes that in the L1 lexical representation, morphological, phonological, orthographic, semantic and syntactic forms of information about a word are strongly integrated with each other, so that access to them is simultaneous when the lexical entry is “opened” (p. 49). By contrast, lack of contextualised target language input may prevent L2 learners from developing these different forms of knowledge. Furthermore, the strong link between the L1 and the existing semantic system means that L1 translation will always feature heavily in L2 vocabulary learning. Hence, Jiang (2000) proposes three stages for the representation and processing of L2 lexical knowledge.

At the initial stage, only the formal information about the target L2 word is represented in learners’ minds, and the processing of the word relies heavily on L1 translation. During the second stage, the “L1 lemma mediation stage”, a link is made between the L2 word and its conceptual representations through the L1 lemma information, albeit weakly, because “the lemma information is copied from L1, rather than created in the process of learning the L2 words, thus not highly integrated into the entry” (Jiang, 2000: 52). Teacher codeswitching to the L1 during vocabulary explanations could,

Vocabulary learning through listening

therefore, aid the development of this second stage by providing a link between the L2 word and its concept. At the final ‘integration’ stage, L2 lemma and lexeme are highly integrated and represented in L2 learners’ mental lexicon.

Jiang (2000) does not however go so far as to advocate L1 use in actual L2 classroom vocabulary instruction (Tian, 2011) and indeed argues that abundant and contextualised L2 input is essential for learners’ L2 lexical representation. Furthermore, he argues that learners may be less motivated to extract further information about a word from contextualised input if they can access its meaning fairly automatically via the L1. Overall, it is unclear how much and what kind of L1 for instructional purpose is beneficial.

The impact of teacher codeswitching within vocabulary instruction after listening was explored in an important study involving 117 university-level learners of English in China by Tian (2011), aspects of which are also reported in Tian and Macaro (2012). Three groups of learners were studied: two intervention groups received vocabulary instruction (‘Lexical Focus-on-Form’) after listening to L2 texts, with vocabulary meaning provided through teacher CS (Chinese to English translation), or through the L2 (explaining word meanings using English only); and a third (control group) was exposed to listening texts without vocabulary instruction, discussing listening strategies instead. The groups receiving intentional vocabulary instruction made significant gains in tests taken at the end of each teaching session. These tests were not administered to the control group on the grounds that it would be inappropriate to assess learners on items for which they had not received instruction, according to Tian and Macaro (2012), although they do accept that their administration might have highlighted short-term vocabulary learning by the control group. Although intervention group scores declined significantly at the delayed post-test (taken by all groups), they were still significantly higher than at pre-test with better long-term vocabulary retention than found in the control group. Vocabulary instruction after listening comprehension activities was also shown to benefit younger L2 learners of French in Hennebry et al. (2013), a rare study in addressing school-level vocabulary learning. They found more effective vocabulary retention through direct teaching than from incidental learning through listening on one of the tests they administered but not on others.

Both Tian and Macaro (2012) and Hennebry et al. (2013) also found that participants who received teacher CS for vocabulary explanations outperformed those who were taught in L2-only. In the first

Vocabulary learning through listening

study, this advantage was found for short-term learning through meaning-recall tests at the end of each session of vocabulary instruction, yet disappeared in delayed post-tests administered two to seven weeks after each instructional session. By contrast, Hennebry et al. (2013) found a longer-term advantage (between one and four weeks after the vocabulary instruction) for teacher CS over L2-only. Effect sizes were however small. Furthermore, in both studies the variation in the length of the gap between instruction and the delayed post-tests makes it difficult to judge how effective the instruction was for long-term learning, suggesting the need for further exploration.

It is possible that more durable and larger gains might be achieved through a third intentional vocabulary teaching method proposed by Laufer and Girsai (2008), in which cross-linguistic information about the target vocabulary is provided. In their study, learners completed text-based translation activities (L1 to L2 and vice versa), but also received information from the teacher that contrasted the L1 and the L2, termed contrastive Form-Focused Instruction. The authors highlight that such instruction does not involve “bilingual glosses which simply state the meaning of L2 words” - namely codeswitching, in which any L1/L2 contrasts remain implicit - but rather “the kind of instruction which leads to learners’ understanding of the similarities and differences between their L1 and L2 in terms of individual words and the overall lexical system”, through the explicit highlighting of such differences (p. 696). Laufer and Girsai (2008) relate contrastive form-focused instruction to Schmidt’s (1990) ‘noticing’ hypothesis, arguing that the contrastive focus can make an L2 word more salient to the learner and hence more likely to be retained. To our knowledge, contrastive form-focused instruction has not been considered within vocabulary development through listening nor compared directly with CS and L2 explanations.

Finally, Tian and Macaro (2012: 371) raise the important question of “optimal L1 use” and highlight the need for research to gain better insights into the “cost-benefits of Lexical Focus-on-Form by the teacher”. In other words, what impact do different kinds of Lexical Focus-on-Form have on other skills, most importantly on listening, the medium through which the instruction is delivered? To our knowledge, this question has not been explored by previous research, which from a pedagogical perspective as well as a theoretical one is an unfortunate omission. A statistically significant relationship between vocabulary and listening has been established in a number of studies, indicating that an approach leading to vocabulary growth might be accompanied by improvements in listening comprehension. Correlations are however typically weaker than those found for vocabulary size and

Vocabulary learning through listening

1
2
3
4 reading, for example, around .70 in Stæhr (2009). These weaker correlations suggest that there is
5 more to listening proficiency than lexical knowledge, as is also suggested in a study by Bonk (2000).
6 He reports that some participants appeared to know under 75% of the words in a spoken text, yet had
7 ‘good’ comprehension of that text. By contrast, others knew over 90% of the words but had ‘poor’
8 comprehension, perhaps suggesting weaknesses in what is called the ‘utilization’ phase of John
9 Anderson’s (2010) model of second language listening, in which perceived linguistic information is
10 related to (non) linguistic information stored in long-term memory in the form of schemata.
11 Increasing learners’ vocabulary knowledge is therefore no guarantee of improved listening
12 comprehension, an argument that also gains support from a systematic review of vocabulary
13 interventions and L1 reading (Wright and Cervetti, 2017). The vast majority of the 36 reviewed
14 vocabulary interventions (either directly teaching vocabulary meanings or teaching a small number
15 of word-deciphering strategies) did not have a positive impact on learners’ general reading
16 comprehension, but only on comprehension of passages that included the vocabulary items taught in
17 the intervention.
18
19
20
21
22
23
24
25
26
27
28
29

30 In sum, there is a gap in research exploring vocabulary acquisition through oral input. The possible
31 pedagogical value of teaching vocabulary through incidental learning plus intentional Lexical Focus-
32 on-Form emerges from the above literature review, yet empirical evidence for this is limited.
33 Furthermore, although using L1 to teach L2 vocabulary is theoretically supported by a
34 psycholinguistic model of vocabulary acquisition and by a small amount of empirical research
35 (notably mainly with university-, rather than school-level learners) no research to date explores
36 specifically which type of L1 explanation is more valuable, namely teacher CS or contrastive (lexical)
37 Focus-on-Form. Finally, even if learners’ vocabulary knowledge can benefit from intentional
38 vocabulary teaching following listening activities, it is unclear what impact such instruction has on
39 other aspects of language learning (notably listening).
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Vocabulary learning through listening

RESEARCH METHODOLOGY

The above issues were explored through two research questions:

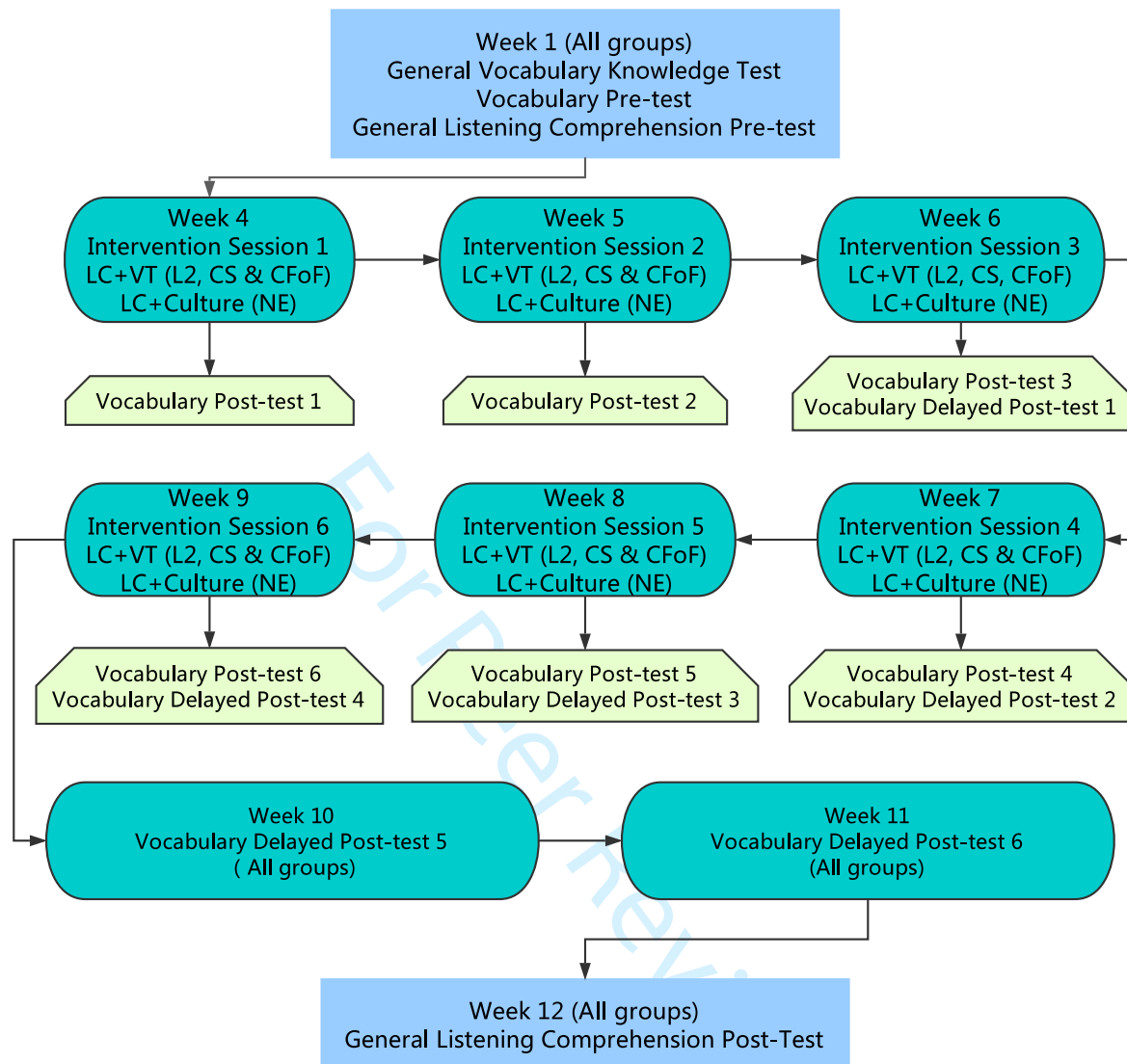
- 1) What is the impact on short and long-term vocabulary learning of: teacher codeswitching; target language explanations; and contrastive Focus-on-Form within vocabulary instruction through listening for high school learners?
- 2) What is the impact of each type of instruction on learners' listening comprehension?

Research Design and Participants

These two research questions were addressed through a quasi-experimental design drawing on Tian and Macaro (2012) and involving 137 first-year high school EFL learners from four intact classes in China from one school (aged 15-16, seven years' English learning experience). Learners were preparing for the Gaokao, China's national university entrance exam, and as such were considered to be of a proficiency level of around A2 to B1 on the CEFR (or around levels 3-4 on IELTSs). The four classes were randomly assigned to three treatment groups, a second language (L2) group, a teacher codeswitching (CS) group and a contrastive Focus-on-Form (CFoF) group, and a No Explanation (NE) group. All groups were instructed by the first author, whose L1 is Chinese.

The study lasted three months, beginning with pre-tests in vocabulary and listening comprehension (week one). Six teaching sessions (outlined below) took place between weeks four to nine. All groups completed an immediate vocabulary post-test at the end of each session. From the third teaching session inclusive (week six), an additional vocabulary delayed post-test was administered at the same time as the vocabulary post-test (six delayed post-tests in total). The delayed post-tests assessed long-term vocabulary retention for target items from the session delivered two weeks previously. For all items there was a delay of two weeks between the post-test and the delayed post-test (comparable with Tian and Macaro (2012) and Hennebry et al. (2013), whose delayed post-tests were administered with a variable delay of between two and seven and two and four weeks respectively). Finally, learners' listening comprehension was reassessed in week 12. Figure 1 depicts the study design. For further details see Author 1 (date).

Vocabulary learning through listening



Note: LC-Listening comprehension; VT-Vocabulary teaching

Figure 1 Outline of all procedures

Research Instruments

Vocabulary At pre-test, a 160 item General English vocabulary test (henceforth, GEVT) was employed in order to gain a measure of general language proficiency as well as to assess knowledge of items to be included in the intervention. The test was based on the aural vocabulary levels test by McLean, Kramer and Beglar (2015) originally for Japanese EFL learners, translated into Chinese and testing only the first three 1000-word frequency levels and the academic word list (as appropriate to the participants' level of language proficiency and ascertained through piloting). Participants heard

Vocabulary learning through listening

the researcher reading out the target lexical item and a neutral example sentence including it. They then needed to select the correct Chinese translation corresponding to the English word, as shown in this example (English translations added here for clarity):

Participants hear: 'School: This is a big school.'

a. 银行 (bank)

b. 海洋动物 (sea animal)

c. 学校 (school)

d. 家 (house)

The 60 target lexical items to be presented in the intervention were intermingled with the 100 items from the general vocabulary test so that the participants would not know which items were the focus of the study. A test using meaning recognition with multiple choice was used at pre-test to facilitate the assessment of partial knowledge of a larger number of items. Cronbach's alpha for this combined vocabulary test was .76, which is in line with what is commonly reported for multiple-choice type vocabulary tests, where test-taking strategies influence reliability to a certain degree (Gyllstad et al., 2015).

Changes in learners' vocabulary knowledge were assessed through six post-tests and six delayed post-test (both tests assessing ten items at a time) based on the assessment used by Tian (2011) but modified so that it took an aural form. The teacher read out the vocabulary items plus an additional sentence which included the target item but did not give contextual clues as to the item's meaning. Learners then had to write down the meaning of each item, with the option of writing either in the L1 or the L2. The reliability for the vocabulary post-test was .94, and .92 for the delayed post-test (Cronbach's Alpha). The post-test and delayed post-test assessed meaning recall without multiple choice to measure growth from the meaning recognition assessed at pre-test. We outline measures taken to address the non-identical format of the tests in the Analysis section.

Listening comprehension At pre-test the first two sections of an IELTS listening practice test were administered. These first two sections include "a conversation between two people set in an everyday social situation" and "a monologue set in an everyday social situation" and as such were considered appropriate for the proficiency level of the participants, with IELTS Band 3-4 assessing

Vocabulary learning through listening

comprehension of familiar situations (British Council, n.d.).

In order to avoid a practice effect a different but equivalent second version of the test was then used after the intervention as a post-test. Steps were taken to ensure that the two tests were of comparable level of difficulty, matching each on 1) vocabulary frequency levels; 2) number of words and speech rate; and 3) the number of global or local questions assessed. Reliability levels (pre-test, .62, post-test, .60) were rather low but consistent with those found in IELTS research reports for the first two sections of the test, namely around .67 (Breeze and Miller, 2012) and similar to levels reported for listening tests used in Stæhr (2009). Out of twenty questions in each test only five were in multiple choice format, in order to limit guessing.

Teaching procedures We followed Tian and Macaro (2012) as far as possible in the implementation of the teaching intervention, with the addition of a Contrastive Focus-on-Form condition. We thus had four groups: No explanation (NE); teacher codeswitching¹ (CS); target language explanations (L2); and contrastive Focus-on-Form (CFoF). The intervention was implemented over six sessions for all four groups, with each session lasting approximately 45 minutes. In each session for all groups learners heard a passage and then answered three questions which asked for general comprehension without focusing on the target items (see Supplemental Materials). More specific comprehension questions were asked orally at this stage in order to again focus learners' attention on listening comprehension.

For the three intervention groups, the passage was then replayed sentence by sentence so that vocabulary explanations of the target items were given as they occurred in the passage, taking a different form for each experimental condition but ensuring that each intervention group received the same amount of vocabulary explanation, as follows (see also Figure 2).

L2 group: Learners firstly received an English explanation for the target item. Then, they were given an additional sentence including the target item with further L2 explanations.

CS group: The meaning of the target lexical item was given by the teacher in the L1. To maintain consistency with the amount of input received by the L2 group, learners were also given an additional L2 sentence including the target item.

Vocabulary learning through listening

CFoF group: Learners were initially given the L1 meaning of the target vocabulary item, then received an additional explanation in the L2 focusing on comparing and contrasting the L2 word and its L1 translation, especially when there was a mismatch between the two.

‘This character was a social failure but was loved for his optimism and determination to *overcome* all difficulties’

L2 explanation: Here, overcome means ‘succeed in a struggle against something’. So ‘overcome all difficulties’ means succeed in a struggle against all difficulties. Another example for this word can be ‘I would overcome any weakness, any despair, any fear.’.

CS explanation: Here, overcome means ‘战胜’. So ‘overcome all difficulties’ means 战胜 all difficulties. Another example for this word can be ‘I would overcome any weakness, any despair, any fear.’.

CFoF explanation: Here, overcome means ‘战胜’ which is equivalent to ‘fight to win’. Therefore, ‘战胜’ can be used to describe people. However, we only use ‘overcome something’ in English. So ‘overcome all difficulties’ means 战胜 all difficulties.

Figure 2 Examples of each type of vocabulary instruction

All intervention groups also saw the written form of the target items, presented through PowerPoint, but they were not allowed to write them down, to prevent any revisiting of the items outside of class. The meaning of the whole listening passage was then summarised in English by the instructor once more and L2, CS and CFoF explanations for the target items were repeated (i.e. two lots of focused sentence encounters were given). Then the whole listening passage was played again. In other words, intervention groups heard the passage three times in total.

The NE group listened to the same passages as the intervention groups and also answered the comprehension questions. Instead of vocabulary explanations, however, they were given cultural and background information (in the L2) relating to the listening passage. To ensure that the NE group had as many focused exposures to the target items as the intervention groups, they were allowed to listen to the listening passage sentence by sentence twice, matching the two rounds of vocabulary explanations received by the treatment groups. Like the intervention groups they heard the passage three times in total and also saw it in written form after the final round of listening.

Vocabulary learning through listening

Target lexical items and listening passages Sixty target lexical items included 18 nouns, 13 verbs, 12 adjectives and 17 collocations. The 43 single words were selected from the 1, 2 and 3K frequency bands and from the academic word list while the 17 collocations were chosen as those included in the senior-secondary school English curriculum for learners at this level (MoE, 2003). The listening passages through which the vocabulary instruction was delivered came from an English text book, 'New Senior English for China'. Passages were selected and modified according to the criteria of: topic relevance to learners; at least 95% of the vocabulary they contained would be familiar to learners (Tian and Macaro, 2012), according to the senior-secondary school English curriculum (MoE, 2003) and the textbook authors; they were of a standard length of approximately 250 words (Tian, 2011); and a uniform speech rate controlled at between 150-190 words/min (Brindley and Slatyer, 2002; Tauroza and Allison, 1990). The topics covered in the passages and the target lexical items did not overlap with those featuring in the listening pre and post-tests.

Research ethics

Informed written consent was obtained from the head teacher, class teachers, parents and students. Issues relating to the appropriateness of assessing NE Group learners in post-tests on items they had not been explicitly taught were considered in consultation with the school and not felt to be problematic from an ethical perspective.

DATA ANALYSIS

Vocabulary pre, post and delayed post-tests were scored by the first author. Post and delayed post-tests were then second marked by another researcher, giving inter-rater reliability rates of between 98.76% and 99.22%. Items were marked either right or wrong, with no half marks. Exploratory factor analysis was firstly carried out on the six post-tests and six delayed post-tests (see Supplemental Materials for descriptive statistics). This indicated only one factor for the post-test (explaining 77.71% of the variance) and one for the delayed post-test (explaining 57.56% of the variance), with relatively high factor loadings (between .58 and .94). Hence it was considered justifiable to aggregate scores for all six post-tests and delayed post-tests, giving one total score for each, following the approach used by Tian and Macaro (2012). For listening, as no half marks were awarded but misspellings were tolerated where these did not impede comprehension or alter the meaning of the answer, all scripts

Vocabulary learning through listening

were marked by the first author and then a second marker double-marked 30% of the pre and post-tests, giving inter-rater reliability rates of 99.34% and 99.56% respectively.

In order to establish comparability between the four groups on linguistic knowledge at baseline, we firstly explored whether they differed significantly on the GEVT. Using a non-parametric Kruskal-Wallis test because of many outliers in the CFoF and the NE group, we established that there were no significant differences between the four groups for vocabulary size ($p = .25$). In order to address the difference in format of the pre-test (meaning recognition, multiple choice) compared with the post and delayed post-test format (meaning recall, no multiple choice), a series of logistic regression tests was performed using generalised linear mixed-effects models (GLMMs) ('lmerTest' package version 3.0-1, R version 3.4.3). These tests took into account the random effects that any test items may have had across different test time points because of the different testing format. Further details are given in the Results section. For all tables, the following applies: * - $p < .05$; ** - $p < .01$, *** - $p < .001$. For effect plots, values on the y axis are given as probabilities, with a maximum of 1. Effect sizes were calculated as odds ratios but were translated into Cohen's d for ease of interpretation, with the latter interpreted as follows: Between-groups contrasts, small = .4, medium = .7, large = 1.0; Within-groups/pre-post contrasts, small = .6, medium = 1.0, large = 1.4 (Plonsky and Oswald, 2014). Where multiple comparisons were made, p -values were adjusted using the 'Hommel' procedure.

RESULTS

Short-term and long-term vocabulary learning

Our first research question explored the impact on short and long-term vocabulary learning of the different types of teacher vocabulary explanations. Descriptive statistics were produced for all groups (Table 1).

Table 1

Descriptive statistics, vocabulary

Group (N)		GEVT	Pre-test	Post-test 1-6 TOTAL	Delayed Post- test 1-6 TOTAL
L2 (35)	Mean (<i>SD</i>)	50.43 (8.06)	10.29 (3.30)	27.80 (8.46)	16.57 (8.00)
	Min–Max	31.00–66.00	4.00–16.00	7.00–46.00	5.00–37.00
CS (36)	Mean (<i>SD</i>)	50.31(6.60)	11.64 (4.85)	44.22 (7.48)	16.53 (8.21)

Vocabulary learning through listening

	Min–Max	35.00-66.00	2.00–21.00	30.00–56.00	6.00–36.00
	Mean (<i>SD</i>)	49.18 (8.69)	6.00 (1.44)	50.00 (6.64)	28.70 (9.16)
CFoF (33)	Min–Max	29.00-72.00	3.00–10.00	29.00–59.00	14.00–49.00
	Mean (<i>SD</i>)	46.94 (7.69)	5.58 (1.73)	9.88 (5.72)	1.67 (2.04)
NE (33)	Min–Max	28.00-61.00	2.00–9.00	1.00–23.00	0.00–9.00

Note. The maximum score for the GEVT is 100.00, for the pre and post-tests, 60.00.

In order to examine whether the participants' vocabulary knowledge varied according to two fixed effects: Time (1 = pre-test, 2 = post-test, 3 = delayed post-test), and Group (L2, CS, CFoF, NE), and their interaction, a series of logistic regression tests was undertaken using GLMMs for each possible Interpretation. NE and Time 1 were set as the baseline level of group and time to be compared with the other groups and test time points. The random factors were Participant and Item (60 items for pre-test, 60 items for post-test and 60 items for delayed post-test; wrong answer coded as 0 while correct answer coded as 1). Random effects were fitted using a maximal random effects structure (Barr et al., 2013). This included random intercepts for Participants and Items, by-participant random slopes for Time, and by-item random slopes for Time, Group, and their interaction.

Final model selection was then carried out followed the “backward selection” procedure (Gries, 2013: 260), whereby a maximal model including the full fixed and random effects structure was fitted initially. Thereafter, while holding the fixed effects structure constant, the random effects structure was simplified by removing the random effect which contributed the least variance. A model comparison between the simplified model and the full model was then conducted. If the simplified model was not significantly different from the full model, the random structure was further simplified by removing another random effect which contributed the least variance, until the point was reached where significant differences were found between the simplified model and the full model. Finally, the fixed effects structure was simplified using the same procedure as for the random effects structure.

For our model for the vocabulary tests, the maximal model was retained as the final model, because removing any random or fixed effect made a significant difference to the model results. The final model represents a good fit to the data: C -value = 0.94, $R^2_{\text{marginal}} = 0.40$ and $R^2_{\text{conditional}} = 0.80$, and there was no significant overdispersion or collinearity (all GVIFs < 3.5). Since in our model, two fixed effects, Group and Time, had more than two categorical levels, several models were run using

Vocabulary learning through listening

different baseline level of Group and Time in order to give a clear picture of each pairwise comparison. Interactions are shown in Table 2, and fixed effects of Time and Group in the Supplemental Materials.

Fixed effects indicate that all intervention groups made statistically significant progress between Time 1 and 2, with large effect sizes. While all intervention groups also significantly improved between Time 1 and 3, effect sizes were large for the CFoF group and approaching medium for the L2 group. There were significant declines between Times 2 and 3 (effect sizes ranging from small for the L2, to medium for the CFoF group and large for the CS group). For the NE Group, statistically significant progress occurred only between Time 1 and Time 2 (medium effect sizes), with a significant decline between Times 2 and 3. The decline between Time 1 and Time 3, although not statistically significant, showed a small effect size.

The Time x Group interactions (Table 2) showed that there were differences in the short-term and long-term progress made by the groups (see also Figure 3).

Table 2

Results from the generalised linear mixed effects model for the vocabulary tests

Interactions	Fixed effects					Random effects
	<i>B</i>	<i>SE</i>	<i>z</i>	<i>p</i>	<i>d</i>	By item <i>SD</i>
Time _{Time2→Time1} x Group _{L2→NE}	1.04	0.62	1.66	.19	0.57	2.54
Time _{Time3→Time1} x Group _{L2→NE}	3.26	0.93	3.51	.005**	1.80	2.63
Time _{Time2→Time3} x Group _{L2→NE}	-2.20	0.71	-3.10	.018*	-1.21	1.71
Time _{Time2→Time1} x Group _{CS→NE}	1.67	0.62	2.73	.043*	0.92	2.62
Time _{Time3→Time1} x Group _{CS→NE}	2.27	0.91	2.50	.061	1.25	2.54
Time _{Time2→Time3} x Group _{CS→NE}	-0.58	0.73	-0.80	.43	-0.32	2.12
Time _{Time2→Time1} x Group _{CFoF→NE}	3.95	0.59	6.72	< .001***	2.18	1.92
Time _{Time3→Time1} x Group _{CFoF→NE}	5.45	0.90	6.05	< .001***	3.00	2.07
Time _{Time2→Time3} x Group _{CFoF→NE}	-1.48	0.73	-2.04	.16	-0.82	2.08
Time _{Time2→Time1} x Group _{CS→L2}	0.63	0.41	1.55	.24	0.35	1.55
Time _{Time3→Time1} x Group _{CS→L2}	-0.98	0.42	-2.34	.096	-0.54	1.31
Time _{Time2→Time3} x Group _{CS→L2}	1.62	0.28	5.70	< .001***	0.89	1.16

Vocabulary learning through listening

$\text{Time}_{\text{Time}2 \rightarrow \text{Time}1} \times \text{Group}_{\text{CFoF} \rightarrow \text{L}2}$	2.91	0.48	6.11	< .001***	1.60	1.90
$\text{Time}_{\text{Time}3 \rightarrow \text{Time}1} \times \text{Group}_{\text{CFoF} \rightarrow \text{L}2}$	2.20	0.50	4.39	< .001***	1.21	2.00
$\text{Time}_{\text{Time}2 \rightarrow \text{Time}3} \times \text{Group}_{\text{CFoF} \rightarrow \text{L}2}$	0.71	0.27	2.61	.054	0.39	0.98
$\text{Time}_{\text{Time}2 \rightarrow \text{Time}1} \times \text{Group}_{\text{CFoF} \rightarrow \text{CS}}$	2.27	0.41	5.59	< .001***	1.25	1.68
$\text{Time}_{\text{Time}3 \rightarrow \text{Time}1} \times \text{Group}_{\text{CFoF} \rightarrow \text{CS}}$	3.18	0.44	7.22	< .001***	1.75	1.94
$\text{Time}_{\text{Time}2 \rightarrow \text{Time}3} \times \text{Group}_{\text{CFoF} \rightarrow \text{CS}}$	-0.91	0.31	-2.91	.029*	-0.50	1.54

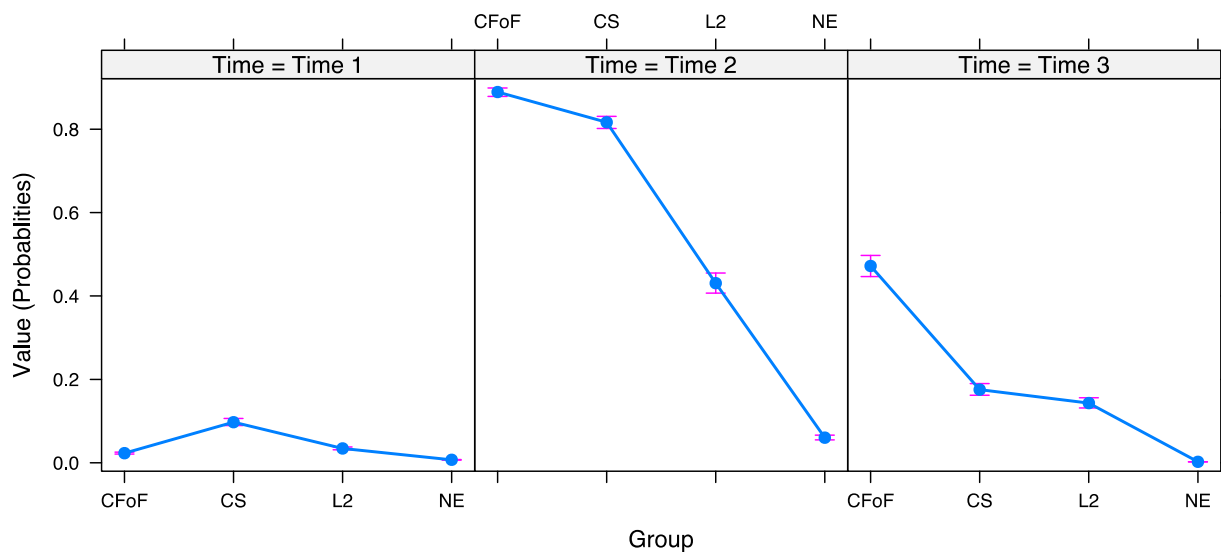


Figure 3 Time*Group effect plot for vocabulary

The CFoF group significantly outperformed the other three groups for both short-term and long-term vocabulary learning, with large effect sizes in all cases. Once random effects had been controlled for, no statistically significant differences were found between the L2 and CS groups for short-term and long-term improvement, but there was an emerging advantage for the CS group in the form of a larger effect size for improvement from Time 1 to Time 2 (2.20 compared with 1.86 for the L2 group – see Supplemental Materials). While the CS group significantly outperformed the NE group for short-term learning (medium effect size), the L2 group was significantly superior to the NE group for long-term learning. Turning to declines in vocabulary knowledge between Time 1 and 2, most notably the L2 group lost significantly less vocabulary knowledge than the NE group (large effect sizes) and the CS group (medium effect sizes).

Vocabulary learning through listening

Listening comprehension

Our second research question investigated the impact of each type of instruction on learners' listening comprehension. Descriptive statistics for the pre and post tests were first calculated (Table 3), removing scores for two learners (L2 and CFoF group) who did not participate in the post-test.

Table 3

Descriptive Statistics (%) for Listening Comprehension

Group (N)		Listening comprehension pre-test	Listening comprehension post-test
L2 (34)	Mean (<i>SD</i>)	21.47 (10.98)	27.94 (8.80)
	Min–Max	5.00-45.00	5.00-40.00
CS (36)	Mean (<i>SD</i>)	17.64 (10.18)	28.19 (11.22)
	Min–Max	0.00-45.00	5.00-50.00
CFoF (32)	Mean (<i>SD</i>)	18.91 (12.81)	21.25 (11.29)
	Min–Max	0.00-55.00	0.00-40.00
NE (33)	Mean (<i>SD</i>)	16.67 (11.43)	37.73 (13.70)
	Min–Max	0.00-40.00	5.00-60.00

In order to analyse whether participants' listening comprehension varied according to three fixed effects -Vocabulary size (GEVT), Time (pre-test vs post-test) and Group (L2, CS, CFoF, NE), and their interaction - a series of logistic regression tests was performed using GLMMs for each possible Interpretation, with NE set as the baseline level of Group and Pre-test set as the baseline level of Time. The random factors were Participant and Item (20 items for pre-test and 20 items for post-test; wrong answer coded as 0 while correct answer coded as 1). Random effects were fit using a maximal random effects structure, which included random intercepts for Participants and Items, by-participant random slopes for Time, and by-item random slopes for Group, GEVT and their interaction. Model selection indicated that the full model could be simplified by removing the fixed effect GEVT from the random effects structure and the Time x Group x GEVT and Group x GEVT interactions from the fixed effects structure. Model comparison indicated that the final simplified model ($AIC = 4326$) did not significantly differ from the full model ($AIC = 4333$), $X^2(6) = 5.04$, $p = .54$.

There was no significant overdispersion or collinearity ($GVI\text{Fs} < 3.5$) for the final model. C -value = 0.90 indicates that the model had a good fit to the data. The two R^2 s ($R^2_{\text{marginal}} = 0.044$, $R^2_{\text{conditional}} =$

Vocabulary learning through listening

0.59), however, indicate that a large part of the variance of the outcome variable was explained by the random effects. We therefore built a null model by removing the fixed effects from the final model and compared our final model with the null model. Results indicate that the final model ($AIC = 4326$) which includes the fixed effects was significantly different from the null model ($AIC = 4342$), $X^2(9) = 33.32, p < .001$. In addition, comparing the model AIC -values shows that the final model was 2119 times better than the null model. Therefore, it was decided to include the fixed effects in the model as they significantly improved the model results.

Model results showed that only the NE group made significant improvement from the pre-test to the post-test, with a medium effect size ($B = 1.75, SE = 0.76, z = 2.29, p = .022^*, d = 0.96$), as well as making significantly greater pre to post-test improvement than the CFoF and the L2 groups did. There was also a main effect of Vocabulary at the Pre-test ($B = 0.045, SE = 0.012, z = 3.75, p < .001^{***}$), indicating that learners' vocabulary size significantly predicted their listening scores. As the fixed effect of Group included four levels, several models were run using each of the four groups as the baseline level of group respectively. Table 4 gives the combined results of all interactions, with effects plots shown in Figures 4 and 5.

Table 4

Results from the generalised linear mixed effects model for listening comprehension

Interactions	<i>B</i>	<i>SE</i>	<i>z</i>	<i>p</i>	<i>d</i>
Time _{Post-test→Pre-test} x Group _{L2→NE}	-1.26	0.48	-2.64	.042*	-0.69
Time _{Post-test→Pre-test} x Group _{CS→NE}	-0.95	0.55	-1.73	.34	-0.52
Time _{Post-test→Pre-test} x Group _{CFoF→NE}	-1.48	0.42	-3.52	.003**	-0.82
Time _{Post-test→Pre-test} x Group _{CS→L2}	0.31	0.37	0.83	.62	0.17
Time _{Post-test→Pre-test} x Group _{CFoF→L2}	-0.22	0.45	-0.50	.62	-0.12
Time _{Post-test→Pre-test} x Group _{CFoF→CS}	-0.54	0.53	-1.01	.62	-0.30
Time _{Post-test→Pre-test} x Vocabulary	-0.028	0.013	-2.16	.030*	/

In addition, the Time x Vocabulary interaction indicates that the lower learners' vocabulary size was at the outset, the more pre-post listening improvement they made.

Vocabulary learning through listening

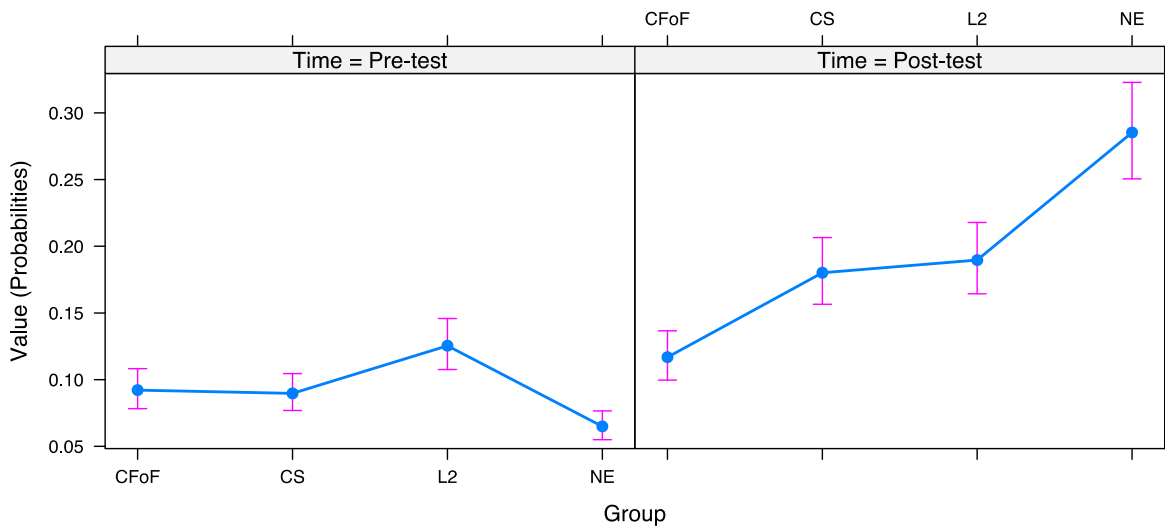


Figure 4 Group * Time effect plot, listening comprehension

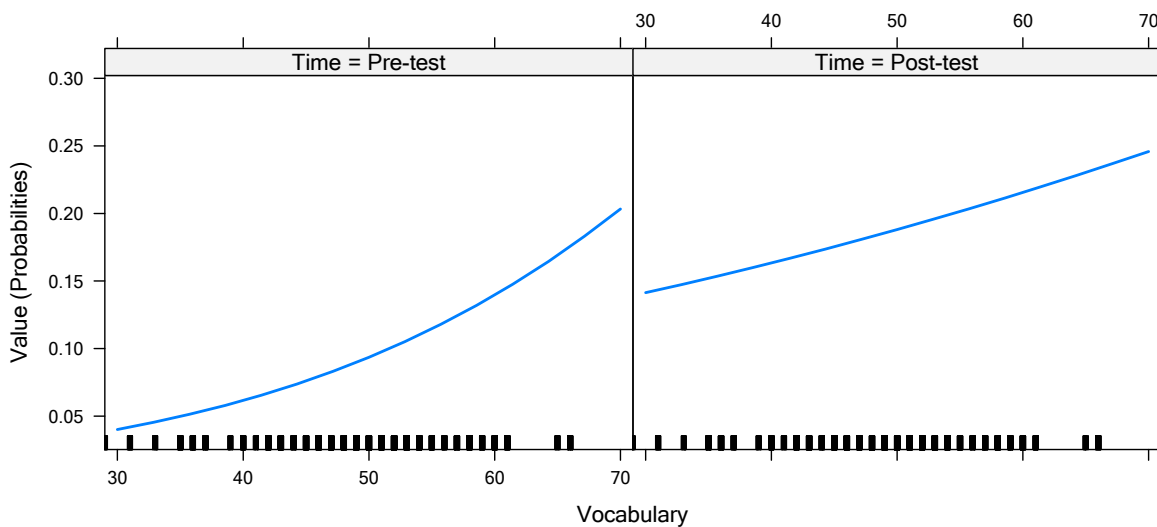


Figure 5 Time * Vocabulary effect plot, listening comprehension

DISCUSSION

This study compared the impact of three types of intentional vocabulary instruction through listening, alongside listening only, on the short-term and long-term vocabulary learning of Chinese EFL learners. An emerging short-term advantage for vocabulary instruction with teacher CS was found compared with L2-only explanations, in terms of greater progress between Time 1 and Time 2,

Vocabulary learning through listening

although not reaching statistical significance. Such a short-term advantage can be interpreted with reference to Jiang's (2000) psycholinguistic model, namely that providing the L1 translation of the target L2 lexical item creates a link between the L2 word and its concept.

Any short-term advantage of the CS group over the L2 group disappeared, however, two weeks after the intervention, a finding similar to that of Tian and Macaro (2012) but differing from the longer-term advantage (1-4 weeks) reported by Hennebry et al. (2013) of using L1 translation over L2-only. In the latter study, however, for some items there was a delay of only one week, with gains made on those items potentially responsible for the overall advantage of the L1 translation group. Furthermore, in the present study, around 36% of the vocabulary knowledge recalled immediately after the intervention was retained two weeks later by the L2 group, compared with only 15% for the CS group. This is similar to what was found by Tian (2011), where learners in the non-codeswitching group retained 23% of the knowledge gained while the CS group only retained 17%. In other words, on balance the evidence suggests that teacher codeswitching for vocabulary explanations does not have long-term benefits when compared with L2-only instruction.

The present study differs from previous studies by exploring a third approach, Contrastive Focus-on-Form for vocabulary instruction. Results suggest both short-term and long-term value for CFoF over both teacher CS and L2, with the highest level of knowledge retention at the delayed post-test (52%). This supports Laufer and Girsai (2008)'s study, where learners who received contrastive form-focused instruction when completing reading comprehension tasks retained significantly more vocabulary items than their counterparts who received L2-only vocabulary explanations at both immediate post-test and one-week delayed post-test. The present study goes one step further however by comparing the giving of cross-linguistic information on vocabulary items with L1 translation as well as L2 vocabulary explanations. The long-term advantage of CFoF over both teacher CS and L2 points to the importance of 'noticing' (Schmidt, 1990), whereby cross-linguistic information makes an L2 word more salient to the learner and hence more likely to be retained.

Additionally, the benefits of teacher CS may be limited because of lexical fossilization, whereby learners do not move from the L1 mediation stage to the L2 integration stage suggested in the theory (Jiang, 2000), potentially because "the presence of L1 lemma within the L2 lexical entry may block the integration of L2 lemma information" (Jiang, 2000:55). Moreover movement to the L2 integration

Vocabulary learning through listening

stage requires learners to extract further information about a word through contextualised input, something which, according to Jiang, may actually be less likely or slower to happen if learners can access the word's meaning fairly automatically via the L1. The fact that learners who received CFoF instruction retained more vocabulary than both the L2 and the CS groups on a short-term as well as long-term basis suggests that CFoF can, to some extent, support L2 learners in moving to, or at least approaching, the L2 integration stage.

Regarding the NE group, although they scored significantly lower than the three treatment groups at post and delayed post-test, they still made small but significant pre to post-test vocabulary gains. This indicates that vocabulary knowledge can be acquired incidentally through listening, confirming what was found by Vidal (2011) among university-level students. Contrary to Vidal (2011) however, incidental learning through listening within the present study was found to be unstable in the long-term. The NE group's performance dropped below pre-test levels at the delayed post-test. This may be attributable to the relatively low language proficiency of these school EFL learners, compared with the university-level participants in Vidal (2011). Like the lowest proficiency students in that study, whose vocabulary gains through listening were small and short-term compared with high proficiency students, learners in the present study may have struggled with speech segmentation, isolating lexical items, inhibiting the establishment of "stable long-term memory representations" (Vidal, 2011: 244-245) in phonological memory that Vidal sees as responsible for vocabulary retention through listening.

Turning to our second research question, and notably to the issue of the "cost-benefits of Lexical Focus-on-Form by the teacher" (Tian and Macaro, 2012: 371), we found a complex relationship between vocabulary and listening comprehension. On the one hand learners' vocabulary size did significantly predict their listening scores, reflecting the strong relationship between these two factors that has been previously established (e.g. Staehr, 2009). Likewise, learners across groups with the lowest vocabulary scores at the outset made the most progress in listening, suggesting that vocabulary instruction helped their listening. On the other hand, the NE group, who made the least progress in vocabulary learning, showed the most marked improvement in listening; likewise the CFoF group, who had the greatest vocabulary gains, did not improve the most in listening scores and indeed were significantly below the NE group on listening at post-test. Such findings are nevertheless consistent with conclusions reported for L1 reading comprehension (Wright and Cervetti, 2017), where

Vocabulary learning through listening

comprehension gains were rarely found to occur where items taught in interventions did not feature in the comprehension tests used, as was the case in the present study. It is also possible that learners in the NE group, receiving no vocabulary explanations, sought to gain a more global understanding of the passages, and had to work harder to make sense of what they heard. While this focus on broader meaning may be less useful for vocabulary acquisition (Vidal, 2011), it may have helped the NE group to develop certain types of strategic behaviour which then enabled them to improve their ability to construct coherent meaning from the listening text.

LIMITATIONS

Random assignment at the individual level was not possible in the study and would also threatened its ecological validity. Establishing comparability between the four groups on the baseline General English Vocabulary Test was undertaken to minimise the limitations posed by non-random assignation, however.

Two further potential limitations concern the non-identical format of the vocabulary pre-test compared with the post- and delayed post-test, and target item selection. How far words are learnt incidentally depends, amongst other things, on how important they are for understanding the meaning of the passage (Wright and Cervetti, 2017) in which they feature. While the majority of target items we selected were important for overall understanding of the intervention passages, we did not control for this factor, which may have had a bearing on levels of incidental learning for the NE group. Both of these two potential limitations were however minimised through the use of generalised linear mixed-effects models.

CONCLUSIONS AND PEDAGOGICAL IMPLICATIONS

Alongside the advantages gained through the use of generalised linear mixed-effects models, the study's strengths also lie in its use of aural rather than written vocabulary tests to bring greater consistency between the testing and teaching modality. Additionally, ensuring an equal testing delay for all vocabulary items, and that the NE group as well as the intervention groups completed all tests (Tian and Macaro, 2012), lends greater credibility regarding the findings relating to long-term and short-term learning. Consequently, we feel able to conclude that within communicative classrooms

Vocabulary learning through listening

in which oral input is a more central component, it is worth promoting vocabulary learning through listening, given that all groups, including the NE group, made at least short-term vocabulary gains. However, vocabulary learning through listening should be supplemented by explicit Focus-on-Form teaching with some, but only short-term, advantage to be gained through code-switched explanations. In order to achieve the greatest vocabulary gains on a long-term basis, however, providing additional information about words through CFoF seems to have particular benefits.

Regarding learners' listening comprehension performance, the CFoF approach, although achieving the most impressive vocabulary gains, was the least beneficial for listening development. By contrast, the NE group made the least progress in vocabulary learning, but showed the most marked improvement in listening. On a pedagogical level, these somewhat contradictory findings suggest that a balanced approach is needed, whereby teachers are clear about what it is they are aiming to achieve from certain classroom approaches. Arguably learners need both opportunities, to focus on listening in its own right, and to experience oral input with Contrastive Focus-on-Form teacher explanations as a way to enhance vocabulary knowledge.

In conclusion, the study not only contributes to our understanding of how vocabulary may be most effectively acquired through listening, but also provides new insights into the impact of such instruction on listening development. In so doing, it enhances understanding of the "cost-benefits of Lexical Focus-on-Form by the teacher" (Tian and Macaro, 2012: 372), thus contributing to an under-researched area, with important implications for classroom practice.

Notes:

1. As pointed out by an anonymous reviewer, in the intervention 'codeswitching' only involved using an L1 equivalent of the target word rather than using both languages for the whole explanation. We recognise that the term 'codeswitching' might be contested here for reasons which space does not permit us to outline, but we have retained it for the sake of simplicity. For a full discussion of codeswitching within classroom contexts see Jin and Cortazzi (2018).

Vocabulary learning through listening

REFERENCES

Author 1 (to be added post-review)

Anderson, JR (2010) *Cognitive Psychology and its Implications* (7th ed). NY, US: Worth.

Barr, D, Levy, R, Scheepers, C and Tily, H (2013) Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68 (3): 255–278.

Bogaards, P and Laufer, B (2004) *Vocabulary in a Second Language*. Amsterdam: John Benjamins Publishing.

Bonk, WJ (2000) Second language lexical knowledge and listening comprehension. *International Journal of Listening* 14(1): 14-31.

Breeze, R and Miller, P (2012) Predictive validity of the IELTS Listening Test as an indicator of student coping ability in Spain. In Taylor L and Weir CJ (eds) *IELTS Collected Papers 2* Cambridge: Cambridge University Press, pp. 487-518.

Brindley, G and Slatyer, H (2002) Exploring task difficulty in ESL listening assessment. *Language Testing* 19(4): 369-394.

British Council (n.d.) IELTS. Available at: <https://takeielts.britishcouncil.org/find-out-about-results/understand-your-ielts-scores> (accessed 02 July 2018).

Gries, S (2013) *Statistics for Linguistics with R: A Practical Introduction* (2nd ed). Berlin: De Gruyter Mouton.

Gu, Y (2018) Types of listening. In Liantas J (ed) *TESOL Encyclopedia of English Language Teaching*. Wiley. <https://doi.org/10.1002/9781118784235.eelt0570>

Gyllstad, H, Vilkaitė, L and Schmitt, N (2015) Assessing vocabulary size through multiple choice formats: Issues with guessing and sampling rates. *ITL-International Journal of Applied Linguistics*, 166(2): 278-306.

Hennebry, M, Rogers, V, Macaro, E and Murphy, V (2013) Direct teaching of vocabulary after listening: Is it worth the effort and what method is best? *Language Learning Journal*. Epub ahead of print 25 November 2013. <https://doi.org/10.1080/09571736.2013.849751>

Hulstijn, JH (2001) Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In Robinson P (ed) *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press, pp. 258-286.

Vocabulary learning through listening

- Jiang, N (2000) Lexical representation and development in a second language. *Applied Linguistics* 21(1): 47-77.
- Jin, L and Cortazzi, M (2018) Use of the first language. In Liantas J (ed) *TESOL Encyclopedia of English Language Teaching*. Wiley.
<https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118784235.eelt0203>
- Laufer, B (2009) Second language vocabulary acquisition from language input and from form-focused activities. *Language Teaching* 42(3): 341-354.
- Laufer, B and Girsai, N (2008) Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics* 29(4): 694-716.
- Lee, JF (2010) Perceptions of ELT among English language teachers in China. *Education Journal* 37(1-2): 137–154.
- McLean, S, Kramer, B and Beglar, D (2015) The creation and validation of a listening vocabulary levels test. *Language Teaching Research* 19(6): 741-760.
- Ministry of Education, P. R. China (2003) *Putong gaozhong yingyu kecheng biao zhun* [English language curriculum guidelines for senior secondary school]. Beijing: People's Education Press.
- Pellicer-Sánchez, A and Schmitt, N (2010) Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language* 22(1): 31–55.
- Plonsky, L and Oswald, FL (2014) How big is “big”? Interpreting effect sizes in L2 research. *Language Learning* 64(4): 878-912.
- Schmidt, R (1990) The role of consciousness in second language learning. *Applied Linguistics* 11(2):17–45.
- Silver, RE, Hu, G and Iino, M (2002) *English Language Education in China, Japan, and Singapore*. Singapore: National Institution of Education, Nanyang Technology University.
- Stæhr, LS (2009) Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition* 31(4): 577– 607.
- Tauroza, S and Allison, D (1990) Speech rates in British English. *Applied Linguistics* 11(1): 90-105.
- Tian, L (2011) *Teacher Codeswitching in a Communicative EFL Context: Measuring the Effects on*

1 Vocabulary learning through listening
2
3

4 *Vocabulary Learning*. Oxford: VDM.
5

6 Tian, L and Macaro, E (2012) Comparing the effect of teacher codeswitching with English-only
7 explanations on the vocabulary acquisition of Chinese university students: A Lexical Focus-
8 on-Form study. *Language Teaching Research* 16(3): 367-391.
9

10 Vidal, K (2011) A comparison of the effects of reading and listening on incidental vocabulary
11 acquisition. *Language Learning* 61(1):219-258.
12

13 Wright, TS and Cervetti, GN (2017) A systematic review of the research on vocabulary instruction
14 that impacts text comprehension. *Reading Research Quarterly* 52(2): 203-226.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

Vocabulary learning through listening

Supplemental material 1

Descriptive statistics for the six subset post-tests and delayed post-tests within each group

	Group (N)	Minimum	Maximum	Mean	SD
	Post1	1.00	9.00	3.97	2.12
	Post2	0.00	4.00	2.00	1.35
	Post3	2.00	9.00	5.66	1.68
	Post4	0.00	9.00	6.71	1.96
	Post5	2.00	9.00	6.26	1.87
L2	Post6	0.00	8.00	3.20	2.22
(35)	Delayed1	0.00	9.00	1.54	1.54
	Delayed2	0.00	4.00	1.94	1.16
	Delayed3	0.00	9.00	2.91	2.39
	Delayed4	0.00	9.00	3.83	2.60
	Delayed5	0.00	9.00	4.09	2.41
	Delayed6	0.00	5.00	2.26	1.12
	Post1	4.00	10.00	7.25	2.18
	Post2	0.00	6.00	3.78	1.61
	Post3	5.00	10.00	8.78	1.46
	Post4	5.00	10.00	8.50	1.61
	Post5	5.00	10.00	8.61	1.38
CS	Post6	2.00	10.00	7.31	2.10
(36)	Delayed1	0.00	10.00	2.44	2.50
	Delayed2	0.00	4.00	2.11	1.06
	Delayed3	0.00	8.00	1.47	2.05
	Delayed4	0.00	9.00	4.00	2.97
	Delayed5	0.00	10.00	4.78	2.64
	Delayed6	0.00	5.00	1.72	1.21
	Post1	4.00	10.00	8.27	1.93
	Post2	2.00	10.00	6.58	1.80
CFoF	Post3	3.00	10.00	9.33	1.61
(33)	Post4	6.00	10.00	9.55	1.03
	Post5	0.00	10.00	8.88	2.42

Vocabulary learning through listening

		Post6	0.00	10.00	7.39	2.69
		Delayed1	0.00	10.00	3.09	3.28
		Delayed2	1.00	6.00	3.58	1.52
		Delayed3	2.00	10.00	6.82	2.35
		Delayed4	0.00	10.00	5.94	2.82
		Delayed5	0.00	10	5.55	3.70
		Delayed6	0.00	10	3.73	3.00
		Post1	0.00	4.00	0.70	1.13
		Post2	0.00	3.00	1.79	0.89
		Post3	0.00	3.00	0.85	1.25
		Post4	0.00	8.00	2.42	2.33
		Post5	0.00	4.00	2.06	1.22
	NE	Post6	0.00	6.00	2.18	1.57
	(33)	Delayed1	0.00	1.00	0.03	0.17
		Delayed2	0.00	3.00	0.48	0.91
		Delayed3	0.00	2.00	0.15	0.44
		Delayed4	0.00	2.00	0.55	0.56
		Delayed5	0.00	2.00	0.21	0.49
		Delayed6	0.00	1.00	0.24	0.44

Vocabulary learning through listening

Supplemental material 2

Sample listening passage, target lexical items and comprehension questions

A Master of Nonverbal Humour

Charlie Chaplin grew more and more popular when he began making films, as his charming character, the little tramp, became known throughout the world. The tramp, a poor, homeless man with a **moustache**, wore large trousers, **worn-out** shoes and a small round black hat. He walked around **stiffly** carrying a walking stick. This character was a social failure but was loved for his optimism and determination to **overcome** all difficulties. He was the underdog who was kind even when others were unkind to him. How did the little tramp make a sad situation entertaining? Here is an example from one of his most famous films, The Gold Rush. It is toward the end of the nineteenth century and gold has just been discovered in Alaska. Like so many others, the little tramp has rushed there in search of gold, but without success. Once he and another man are hiding in a small **hut** during a snowstorm with nothing to eat. They are so hungry that the little tramp tries boiling one of his **leather** shoes for dinner. The little tramp **cuts off** the leather top of the shoe and shares the shoe with the other fellow. He tries cutting and **chewing** the bottom of the shoe as if it were the finest steak. Then he **picks out** the lace of the shoe and eats it as if it were spaghetti. He eats each mouthful with great enjoyment. The acting is so **convincing** that it makes you believe that it is one of the best meals he has ever tasted!

Listening comprehension questions

1. **Why was the character, the little tramp, loved by so many people?**

- A. Because the character is played by Charlie Chaplin.
- B. Because of his optimism and determination.
- C. Because of his funny dressing style.

2. **Why did the little tramp rush to Alaska at the end of the 19th century?**

- A. To search for gold
- B. To change the sad situation that he was suffering
- C. To search for success

3. **What did the little tramp and his friend eat for dinner during a snow storm?**

- A. Spaghetti
- B. Steak
- C. Leather shoes

Vocabulary learning through listening

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

Vocabulary learning through listening

Supplemental material 3

Model results for the main effects of Time (Vocabulary tests)

	Main effects	<i>B</i>	<i>SE</i>	<i>z</i>	<i>p</i>	<i>d</i>
L2	Time _{Time2→Time1}	3.37	0.53	6.35	< .001***	1.86
	Time _{Time3→Time1}	1.77	0.53	3.36	< .001***	0.98
	Time _{Time2→Time3}	1.60	0.23	6.99	< .001***	0.88
CS	Time _{Time2→Time1}	3.99	0.38	10.57	< .001***	2.20
	Time _{Time3→Time1}	0.78	0.37	-2.11	.035*	0.43
	Time _{Time2→Time3}	3.22	0.26	12.56	< .001***	1.78
CFoF	Time _{Time2→Time1}	6.28	0.46	13.58	< .001***	3.46
	Time _{Time3→Time1}	3.97	0.44	8.93	< .001***	2.19
	Time _{Time2→Time3}	2.31	0.22	10.42	< .001***	1.27
NE	Time _{Time2→Time1}	2.33	0.65	3.60	< .001***	1.28
	Time _{Time3→Time1}	-1.49	0.95	-1.57	.23	-0.82
	Time _{Time2→Time3}	3.80	0.69	5.48	< .001***	2.10

Vocabulary learning through listening

Model results for the main effects of Group (Vocabulary tests)

	Main effects	<i>B</i>	<i>SE</i>	<i>z</i>	<i>p</i>	<i>d</i>
Time 1	Group L2→NE	1.66	0.50	3.30	.005**	0.91
	Group CS→NE	2.91	0.52	5.55	< .001***	1.60
	Group CFoF→NE	1.25	0.49	2.52	.056	0.69
	Group CS→L2	1.25	0.35	3.50	.003**	0.68
	Group CFoF→L2	-0.42	0.39	-1.09	.70	-0.23
	Group CFoF→CS	-1.66	0.34	-4.93	< .001***	-0.92
Time 2	Group L2→NE	2.70	0.40	6.77	< .001***	1.49
	Group CS→NE	4.58	0.39	11.62	< .001***	2.52
	Group CFoF→NE	5.19	0.41	12.73	< .001***	2.86
	Group CS→L2	1.87	0.33	5.72	< .001***	1.03
	Group CFoF→L2	2.49	0.33	7.55	< .001***	1.38
	Group CFoF→CS	0.62	0.30	2.07	.16	0.34
Time 3	Group L2→NE	4.92	0.80	6.18	< .001***	2.71
	Group CS→NE	5.17	0.77	6.75	< .001***	2.85
	Group CFoF→NE	6.70	0.79	8.51	< .001***	3.69
	Group CS→L2	0.25	0.34	0.74	.88	0.14
	Group CFoF→L2	1.79	0.35	5.05	< .001***	0.98
	Group CFoF→CS	1.53	0.33	4.56	< .001***	0.84