The Effects of Varying Practice Modes on L2 Learners' Vocabulary Retention Through Songs: Listening, singing, and oral reading

Anna C-S Chang
Hsing Wu University
<annachang@livemail.tw>

Abstract

Singing or listening to songs is one of the most common L2 activities outside the classroom, and many language teachers integrate songs in their teaching; however, the effect of varying song practice modes on L2 learners' retention of linguistic knowledge is under-researched. Eighty-seven EFL learners of university students were recruited and divided into three subgroups. To avoid different song different effect, the study adopted a counterbalanced designed, in which each subgroup practiced three songs under three modes: listening + singing, listening + reading lyrics orally, and listening only. Three aspects of vocabulary knowledge (aural and written meaning recognition, written form production) were measured in the pre-test, one-week delayed, and eight-week delayed post-tests. HLM was performed to analyze the data. Results showed that the main effects of the four variables (times, practice modes, vocabulary dimensions, songs) had significant effects on L2 learners' vocabulary retention, but the interaction effects between times and the three other variables were less salient. Overall, students scored higher for singing mode and for written meaning recognition; however, different songs and times seem to have played an important but complicated role on vocabulary retention. Applications of integrating songs in teaching L2 are discussed.

Keywords: Practice modes, vocabulary retention, songs, singing, oral reading, listening

This study aimed to investigate the effects of varying practice modes of songs on EFL learners' retention of multidimensional knowledge in vocabulary. Songs have been found to be the most common type of L2 exposure outside the classroom (Lindgren & Munoz, 2013) for L2 learners, and a large number of language teachers also integrated songs in their language teaching (see Tegge, 2018). Given this, songs seem to be another valuable vehicle for L2 learners to gain massive aural input to improve their listening skills and vocabulary knowledge. Song lyrics also contain many linguistic features (see shortly below), such as the use of a large proportion of highly frequent vocabulary words and in conversational style, which may be considered easier and more practical for lower-level students. Moreover, people tend to repeatedly sing their favorite songs; the repetition may increase the opportunity for exposure to the language and thus enhance the learning

rates, in particular for vocabulary retention. Despite the abovementioned observations, studies into song effects on L2 learners are limited, and it is particularly so for adult learners.

Literature Review

In the following, the linguistic features of song lyrics based on some corpus studies will be first examined, followed by reviewing some empirical and relevant studies.

Theoretical Benefits of Using Songs as Learning Materials

Theoretical benefits of integrating songs into language teaching have shown that song lyrics contain some language features that make songs valuable learning materials. The first feature is that songs comprise most frequent vocabulary words. Some larger corpus studies into lexical demand of pop songs show that more than 95% of the lexical coverage was from the first 3,000-word levels. For example, Tegge (2017) studied 408 English-language pop songs from 2008 through 2014 and found that the first 3,000 most frequent word families make up 88.96%, 4.22%, and 1.95% of the running words. Hirata (2018) also analyzed 208 songs and found that the first 3,000 words cover 96.11% in the lyrics' corpus. The accumulative coverage of the first 3,000 words was quite similar between Tegge and Hirata although Tegge collected 200 more songs than Hirata.

The second linguistic feature is that song lyrics are highly repetitive. Repeated listening to songs may potentially increase listening comprehension and vocabulary learning. Hirata (2018) found that more than 45% of the word families from the 1,000 – 3,000 level occurred five or more times. When people sing songs they like, they tend to sing them over and over again. This means that the number of encounters with certain words is multiplied. The third feature regards song lyrics containing many short texts and non-specific referents e.g., you, I, he, and she. According to Murphey (1992), 25% of the sentences in his corpus were imperatives and questions, which means the discourse of songs are like authentic conversation and might be suitable for lower-level learners. These above linguistic features may explain why songs are widely considered a valuable teaching source.

Empirical Evidence of Song Effects on Language Teaching and Learning

Different input or practice modes seem to result in different learning outcomes regardless of L1 or L2 (e.g., Moradi & Shahrokhi, 2014; Schön et al., 2008). The latter compared the effects between speech sequences and sung sequences on native speakers of French, and found the performance of the group in the sung sequences was superior to the group in the speech sequences, especially for word segmentation. Similar results were reported by Moradi and Shahrokhi (2014) with two groups of 9-to-12-year-old Iranian female students. The experimental group, who learned and sang songs, achieved significantly higher scores on the tests of pronunciation of segmental sounds than the control group, who learned the same songs through spoken texts without musical accompaniment.

Some studies investigated whether singing supports vocabulary learning better than other approaches (e.g., choral repetition, games, and stories) for very young children (Coyle & Gracia, 2014; David & Fan, 2016). For example, Coyle and Gracia (2014) explored the effects of song-based activities on the acquisition of vocabulary by a group of 25 preschool Spanish children learning English as foreign language. The children listened to songs seven times during three 30-minute class sessions. Students were given vocabulary picture tests before and immediately after the three lessons, and a five-week delayed-post after the last sessions. The researchers found that

using a song to teach English to preschoolers helped the majority of the children develop their receptive knowledge of vocabulary and but not productive knowledge. The major limitation of this study is that the intervention time lasted only 90 minutes in total. The exposure time might not be sufficient to see the effect for productive knowledge to develop. Another possible reason that no productive knowledge developed could be that the children were too young to produce what they had learned.

More recently, two studies extended the age group to include elementary pupils (Chou, 2014; Pavia, 2023; Pavia et al., 2019). For example, Pavia et al. (2019) investigated 300 EFL Thai students in grades five and six. Students listened to two songs (A & B) once, three times, five times, or no listening at all, and were tested on their learning on spoken-form recognition, form-meaning connection, and collocation recognition immediately after they had listened to each song and also a week after the exposure. Mixed results were shown, making it difficult to tease out the tangle of different aspects of vocabulary knowledge, the number of repetitions, different songs, and different times. Pavia (2023), however, further explored the role of different modes of input with another group of Thai students learning formulaic sequences. Four input modes were compared: listening only, listening while reading the lyrics, listening and singing, and listening and singing while reading the lyrics, and the last input mode was found to be most effective in gaining formulaic sequences.

The above studies seem to focus on children. A recent study on young adults was conducted by Baills, Zhang, Cheng, Bu, and Prieto (2021). Baills et al. (2021) found no significant differences in word recall for Chinese learners of French between song singing and rhythmic speech, nor was it between song singing and song listening. Because the intervention lasted only four minutes with three repetitions, students' might have been able to pay their full attention to the input, which could make it difficult to distinguish whether one practice mode was better than the other. The two studies did not involve a delayed post-test, so it is unclear whether singing songs has a long-lasting effect for adult learners.

Taken all together, the findings from the above literature review are synthesized below in terms of four fixed variables related to the present study:

1. Practice modes: There is no conclusive finding as to whether one mode is superior to another; however, singing was found to be more effective than other practice modes in vocabulary learning in the studies by Schön et al. (2008) and Pavia (2023), no difference in Baills et al. (2021), Chou (2014), and Davis and Fan (2016).

2. Vocabulary dimensions: Singing better facilitates the acquisition of receptive knowledge than productive knowledge (Coyle & Gracia, 2014). Sung sequence is also more helpful for children's word segment and pronunciation than spoken sequences. Pavia et al. (2019) found listening to songs facilitated students' form-meaning connection better than spoken-form recognition and collocation recognition; however, listening plus singing and reading song lyrics yield even better effects (Pavia, 2023).

3. Different songs: Different songs with different vocabulary levels may lead to different gains in vocabulary knowledge (Pavia et al., 2019).

4. Time effects: Most studies found L2 learners, regardless of their ages, could acquire vocabulary knowledge immediately after exposures to songs, but most studies did not involve delayed post-tests (e.g., Baills et al., 2021). Mixed results were found in Pavia et al. (2019).

Coyle and Gracia (2014), however, found students scored much higher for receptive knowledge in the delayed post-test than in the immediate post-test, and productive knowledge did not deteriorate in the five-week delayed post-test.

The Present Study

From the literature review, the song studies on adult L2 learners were limited. The two studies reviewed above had very short intervention periods, 4 and 15 minutes only, and the results were inconsistent. The present study, hence, focused on the retention of three dimensions of single-word items (SWI) at three time points with three different songs. Each student experienced three different learning conditions: listening plus singing, listening plus oral reading, and listening only. Listening only is considered the baseline of learning vocabulary through songs. Two research questions were addressed:

1. To what extent is the retention of SWI through songs affected by time, practice modes, different songs, and vocabulary dimensions?

2. Were there any significant interactive effects between time and practice modes, songs, and vocabulary dimensions?

Method

Participants

This study involved 120 technological university students majoring in varying subjects, such as hospitality, tourism, and media, among whom 87 completed the full intervention. The students' native language was Chinese, and they enrolled in a required English course. The time for the course was 100 minutes per week. The students were formed into three subgroups (A, B, and C) by the university. A placement test administered by the university language learning showed that their English proficiency ranged from high beginner to low intermediate; subgroup A (n = 29) scored the highest, followed by subgroups B (n = 33), and C (n =25). Due to the difference in their language proficiency (LP), their LP was used as a covariate in data analysis.

Research Design and Procedure

The three subgroups were told to study three songs using a different practice mode, which was determined by a draw. Each subgroup received three different types of treatment each week: listening plus singing, listening plus oral reading, and listening only (See Table 1). In total, each week every student practiced all three songs, each in a different mode. Before starting the intervention, the students wrote down the lyrics of the whole songs, and the instructor explained the lyrics to help students to comprehend the meanings of the songs. The instructor also explained the meanings of some difficult words, but she did not particularly focus on the target items. During the intervention, students did the following tasks for six weeks and all three groups did the three activities in the same sequence (singing, reading aloud, or listening). The details are described below:

The listening plus singing intervention. Each week the students read the lyrics while listening and singing along with the target song for 10 minutes (about two times), making a total of 60 minutes for the practice time. After six weeks, they video-recorded their singing on their cell phones and sent the recordings to their teacher.

The listening plus oral reading intervention. Each week, the students spent 10 minutes reading the lyrics while listening to the target song once and then reading aloud the lyrics they had written.

After the end of the intervention, students had to video-record their oral reading of the song on their cell phones and send the recordings to their teacher.

The listening only intervention. Each week, students read the lyrics while listening to the target songs for about 10 minutes. The total time for listening only was 60 minutes. It has to be noted that this condition is seen as the baseline of learning a language through songs.

Table 1. Summary of the Research Design in the Three Intervention Conditions

	Listening plus singing along	Listening plus oral reading	Listening only
Subgroup A	Song 1	Song 2	Song 3
Subgroup B	Song 2	Song 3	Song 1
Subgroup C	Song 3	Song 1	Song 2

Note. Song 1: "The Sound of Silence;" Song 2: "Starry, Starry Night;" Song 3: "The Colors of the Wind"

Selected Songs and Lyrics Analysis

The criteria for selecting the target songs were as follows: 1) The selected songs had to contain a certain number of words that were unfamiliar to the students; 2) The songs, to a very large extent, were unfamiliar to them; 3) The song lyrics do not contain vulgar expressions. After discussion with three English teachers and senior students, three songs were finalized. They were "The Sound of Silence," composed by Paul Simon and sung by Paul Simon and Art Garfunkel, "Starry, Starry Night," composed and sung by Don McLean, and "The Colors of the Wind," composed by Alan Menken and Stephen Schwartz, and sung by Vanessa Williams. The lyrics were analyzed using BNC-COCA-25000 (downloadable from https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs). The vocabulary frequency levels of each song are shown in Table 2. As can be seen, the word types for the three songs were quite comparable, with 124, 145, and 124 respectively for the three songs.

Table 2. The Vocabulary Frequency Levels for the Three Song Lyrics

Word levels	TOKENS/%			TYPES/%			FAMILIES		
	Sound of Silence	Starry Night	Colors of Wind	Sound of Silence	Starry Night	Colors of Wind	Sound of Silence	Starry Night	Colors of Wind
One	177/81.19	269/88.20	295/91 .05	90/72.58	115/79.31	101/81.45	82	100	91
Two	21/9.63	8/2.62	14/4.32	16/12.90	6/4.14	10/8.06	15	6	9
Three	8/3.67	5/1.64	1/0.31	7/5.65	5/3.45	1/0.81	7	5	1
Four	3/1.38	7/2.30	4/1.23	3/2.42	5/3.45	2/1.61	3	5	2
Five	1/0.46	3/0.98	0/0.00	1/0.81	3/2.07	0/0.00	1	3	0
Six	0/0.00	6/1.97	1/0.31	0/0.00	5/3.45	1/0.81	0	5	1
Seven	3/1.38	2/0.66	2/0.62	2/1.61	2/1.38	2/1.61	2	2	2
Eight	1/0.46	0/0.00	1/0.31	1/0.81	0/0.00	1/0.81	1	0	1
Nine	1/0.46	1/0.33	0/0	1/0.81	1/0.69	0/0	1	1	0
Ten	0/0	0/0	1/0.31	0/0/	0/0	1/0.81	0/0	0/0	1
Proper nouns	0/0	0/0	1/0.31	0/0	0/0	0/0.0	0/0	0/0	1
Not in the lists	3/1.38	4/1.31	4/1.23	3/ 2.42	3/2.07	4/3.23	?	?	?
Total	218	305	324	124	145	124	112	127	109

Note. Family here refers to a word family. A word family includes a word's base forms (e.g., act), its inflected forms (e.g., acting), and its derived forms (e.g., action) (Nation, 2013).

Target Items

Sixty target words were selected from the three songs, among which 20 were from "The Sound of Silence," 27 from "Starry, Starry Night," and 13 from "The Colors of the Wind." As shown in Table 3, more than a half (34/60) words were selected from the first 3,000-word levels, and the rest ranged from the 4,000-to 9,000-word levels. The criteria for selecting these words were as follows: these words must appear in the song lyrics, and no words were fully familiar to the students. For example, students may be familiar with the pronunciation and meaning of "bow" when it is used as a noun, but in the song, "bow" is used a verb and pronounced differently. Overall, most of the selected words were fully unfamiliar to the student participants. Times of repetition of the target words in the songs was not considered a criterion in this study because songs were short and could be easily repeated.

Table 3. Profile for the 60 Target Words

Word levels	Total (%)	Target words
1,000	6 /10%	beneath, bloody, hidden, restless, weathered, worth
2,000	18 /30%	bow, claim, connect, creature, creep, disturb, flaming, flash, grin, narrow, pine, pray, remain, shadow, soul, spirit, split, suffer
3,000	10 / 16.67%	chill, crush, damp, echo, grain, portrait, reflect, trail, vision, whisper
4,000	8 / 13.33%	blaze, breeze, collar, copper, sanity, sketch, stab, virgin
5,000	4 / 6.67%	prophet, soothe, swirling
6,000	5 / 8.33%	hue, palette, ragged, thorn
7,000	5/ 8.33%	haze, heron, hoop, neon, violet
8,000	2/3.33%	halo, otter
9,000	2/3.33%	daffodils, tenement
Total	60/100%	

Dependent Measures

The dependent measures involved two forms of vocabulary test to measure students' vocabulary knowledge in three dimensions, a 19-item questionnaire to explore students' perceptions of learning English through songs (data were not reported in this paper but the questionnaire can be seen in the appendix), and post-test interviews. For the vocabulary test, Form A was to test aural meaning recognition (AMR) and written form production (WFP), and Form B, written meaning recognition (WMR). Each aspect is described below.

Aural meaning recognition (AMR) and written form production (WFP): Students heard a sentence, and then the target word to be tested was repeated once. The students had to select one correct meaning from six options and then write the spelling of the target word. For example, students heard a sentence: "Everyone bowed down before him." Students then had to select one of the correct meanings from six options:

a. 繫蝴蝶結 b. 跪下 c. 擁抱 d. 親吻 e. 鞠躬 f. 後退, and write b-o-w-e-d on the sheet.

Each sentence contains no clues that allow students to guess the meaning.

Written meaning recognition (WMR): In this test, students had to select an equivalent Chinese meaning for each of the underlined words. The meaning of the target item must correspond to the song context. One example is given below:

Everyone bowed down before him.

a.　　跪下 b. 後退 c. 鞠躬 d. 親吻 e. 擁抱 f. 繫蝴蝶結

Individual Interviews

Twelve students were invited to talk to the researcher before the eight-week delayed post-test. Their reflections were used to explain the test results. The interview in particular focused on what students did between the one-week post-test and eight-week delayed post-test. This was to examine whether students continued singing or listening to the target songs after the intervention. The main questions were as follows:

• How well did you like the songs we practiced in class?

• Did you continue singing or listening to the target songs from time to time?

• Did you find it very anxiety-provoking to record your singing or oral reading and present it to your teacher?

• What do you think you learned from these songs?

Intervention

All students were given a pre-test on the three dimensions of the target words, and then were given three sheets to write down the complete lyrics for the three songs, followed by the instructor explaining the meanings of some vocabulary items and the backgrounds of the target songs and performers. For example, who is Don Mclean, when and why did he write the song, and for whom did he write the song? Each week, students took out the worksheets for practice. They practiced the three songs for 30 minutes, approximately10 minutes for each song in different practice modes: listening plus singing, listening only, listening plus orally reading the lyrics. After six weeks, the songs they had to sing and read orally were video-recorded on their cell phones and sent to their teacher. This course of action was to ensure that the students actually did the practice according to the teacher's request. The first post-test was administered to all students one-week after the intervention, and the second post-test was administered eight weeks after the first post-test. The consent forms were signed by the students after the researcher explained the teaching and research purpose. The research procedure is summarized in Table 4.

Table 4. Summary of the Research Procedure

Tasks by Week

Week	Tasks
1	Administering the pre-test of all measures, filling out the before-intervention questionnaire, and explaining the meanings of the song lyrics.
2-7	Intervention period.
8	Administering one-week delayed post-test of all measures and filling out the post-intervention questionnaire.
9-15	Individual interviews.
16	Administering the eight-week delayed post-test of all measures. Students signing the consent form for researcher using their test results for the research purpose.

Scoring and Data Analysis

The multiple-choice test questions were marked by two teaching assistants, and the written form production was first marked by two English teachers. Students' answers to each item were first entered as 0 or 1 (0 for an incorrect answer and 1 for a correct answer), which resulted in a total of 2,349 cases. After summing up the total number of correct items, the scores were transformed into percentages (due to the numbers of words selected from each song being different). SPSS version 25 for Windows was used to analyze the data. To answer the two research questions, the hierarchical linear modeling (HLM) was performed. The reason for using HLM was that it allows all data to be analyzed at one time. The data set in this study is a two-level nested structure (inter-individual level and intra-individual level). More reasons for the use of HLM can refer to Huta (2014). In this study, the dependent variables were students' test scores. The random factor was the student participants, and the fixed factors involved Time (3 levels. 1: pre-, 2: one-week delayed post-, 3: eight-week delayed post- tests), Practice modes (3 levels. L: listening, O: oral reading; S: singing), Vocabulary dimensions (3 levels: AMR: aural meaning recognition; WFP: written form production; WMR: written meaning recognition), Songs (three levels. 1: "The Sound of Silence;" 2: "Starry, Stary Night;" 3: "The Colors of the Wind"). Due to differences in the students' language proficiency, students' language proficiency was used as a covariate.

Results

Descriptive Statistics

The descriptive statistics (in percentage) for students' test scores are presented in Table 5. As shown, students scored comparably for different practice modes, and much higher in the WMR dimension, but much lower for Song 3. From Time 1 to Time 2, and to Time 3, students made the most advancement in singing mode, in the dimension of WMR and in Song 3. Interestingly, students continued to make some advancement after the treatment had ended. This can be seen in that students gained more knowledge between Time 2 and Time 3. Overall, students scored an average of 21.91% of the pre-test vocabulary knowledge of the three target songs. After 180 minutes of practice, they gained 9.92%, but the gains did not come to an end after the intervention. Their knowledge kept on growing. This phenomenon was interesting and will be discussed shortly.

Table 5. Descriptive Statistics for Students' Performance (in %; Maximum Score = 100)

Fixed variables	Levels	Time 1	Time 2	Time 3	Time 1-Time 2	Time 1-Time 3	Time 2-Time 3
Practice modes	Listen	21.78	29.81	34.58	8.04	12.80	4.77
	Oral	21.69	30.81	34.82	9.12	13.13	4.01
	Sing	24.69	34.31	38.44	9.61	13.75	4.13
Vocabulary dimensions	AMR	27.80	35.48	42.25	7.68	14.45	6.77
	WMR	33.99	43.99	49.54	10.00	15.55	5.55
	WFP	6.37	15.47	16.06	9.10	9.69	0.59
Songs	Song 1	27.83	35.24	39.47	7.40	11.64	4.23
	Song 2	22.24	31.45	36.04	9.21	13.80	4.58
	Song 3	18.09	28.24	32.34	10.15	14.25	4.09
Total		21.91	31.83	36.12	9.92	14.21	4.29

Inferential Statistics

Table 6 presents the analysis results of HLM. The interaction effects were first examined but showed no salient effects in most of the terms, so the focus will be on the main effects. In terms of Time, no significant difference was observed between Time 1 and Time 2 ($B = 5.27$, $p = .058$); however, a significant difference was found between Time 1 and Time 3 ($B = 12.43$, $p < .001$). The main effect of practice modes shows that the S mode produced better effect than O mode ($B = 3.00$, $p < .001$), and also better than L mode ($B = 2.92$, $p < .001$), and all the interaction terms between practice modes and Time were nonsignificant, which meant that the effect of input mode was not affected by Time, and S mode is found to be consistently more effective than L and O modes (see Figure 1).

The significant interaction effects between Dimensions and Times were found for "WFP vs WMR" from Time 1 to Time 2 ($B = -5.85$, $p = .046$), indicating that from the pre-test to the delayed post-test, students scored better on WMR than on WFP. Considering the very weak interaction effects, let us look at the main effects of vocabulary dimensions alone. Statistically significant main effects of dimensions were found for WMR vs AMR, ($B = 6.19$, $p < .001$), indicating students performing better on written meaning recognition than aural meaning recognition. Both WMR and AMR were scored significantly higher than WFP, WFP vs WMR ($B = -27.63$, $p < .001$), and WFP vs AMR ($B = 21.44$, $p < .001$) (also see Figure 2). Productive knowledge is always more difficult to acquire than recognitive knowledge. The main effects of Songs show that students consistently scored significantly higher for Song 1, followed by Song 2 and Song 3 (See Figure 3); however, if Time is taken into account, the interaction effect between Songs and Time were found significant for "3 vs 1" (T2-T1)" ($B = 2.75$, $p = .024$) and "3 vs 1" (T3-T1)" ($B = 2.61$, $p = .026$). The results mean that from T1 to T2 and T1 to T3, students gained more vocabulary knowledge from Song 3 than from Song 1 (see Figure 3).
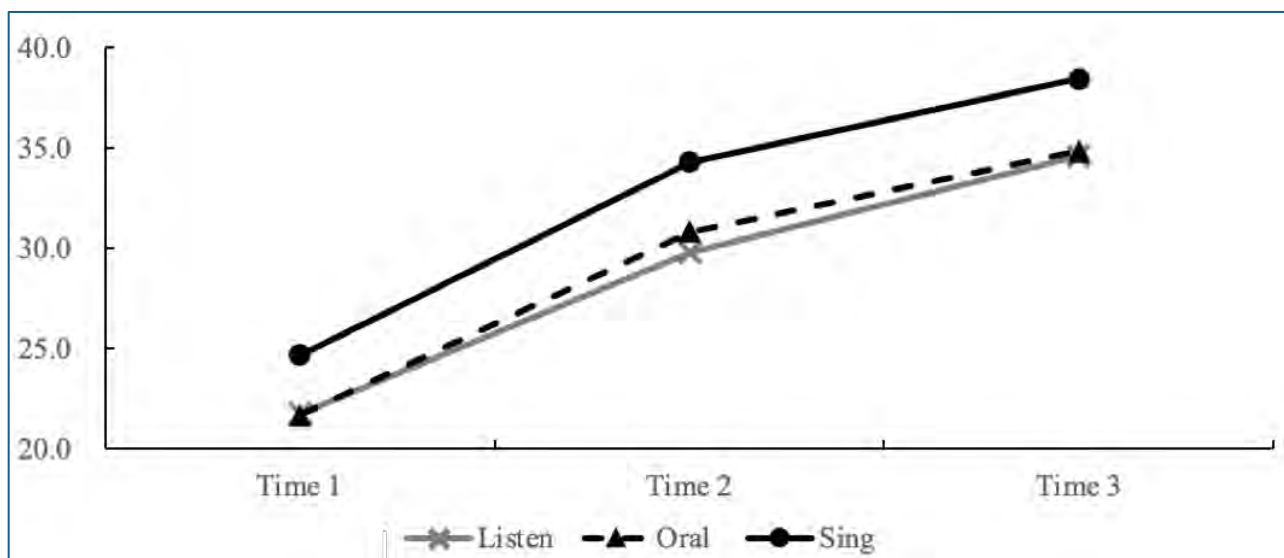

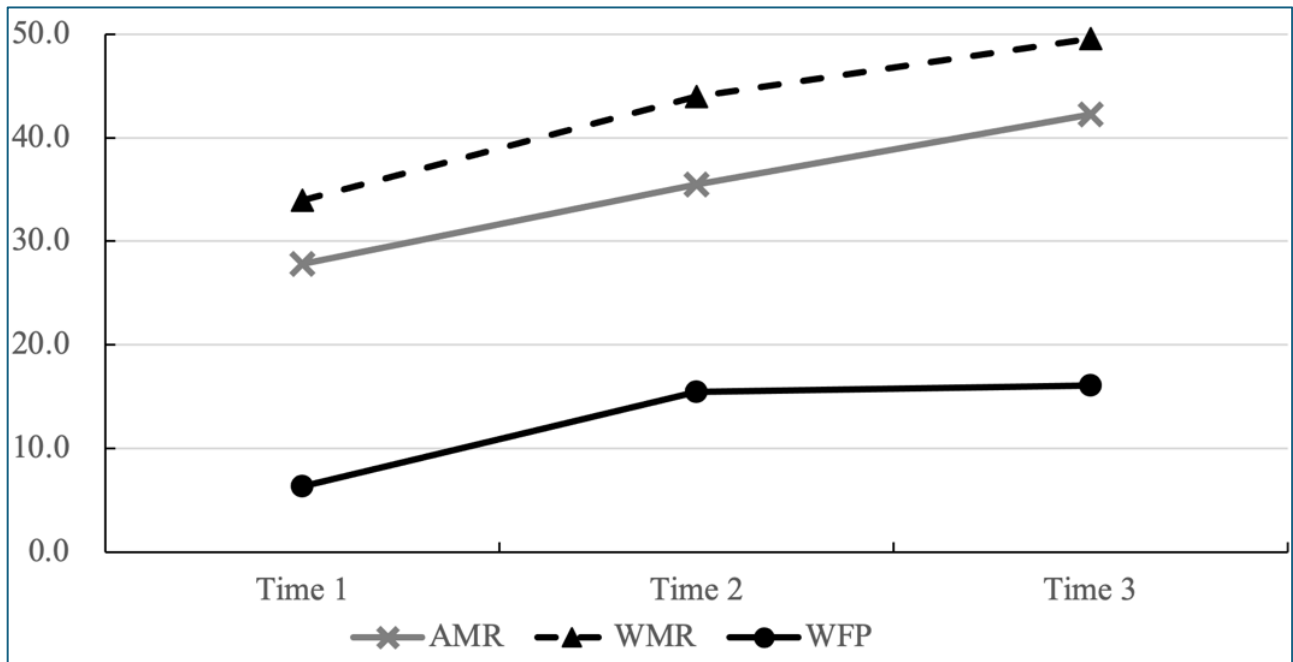**Figure 1. The Interaction Effects Between Input Modes and Times**

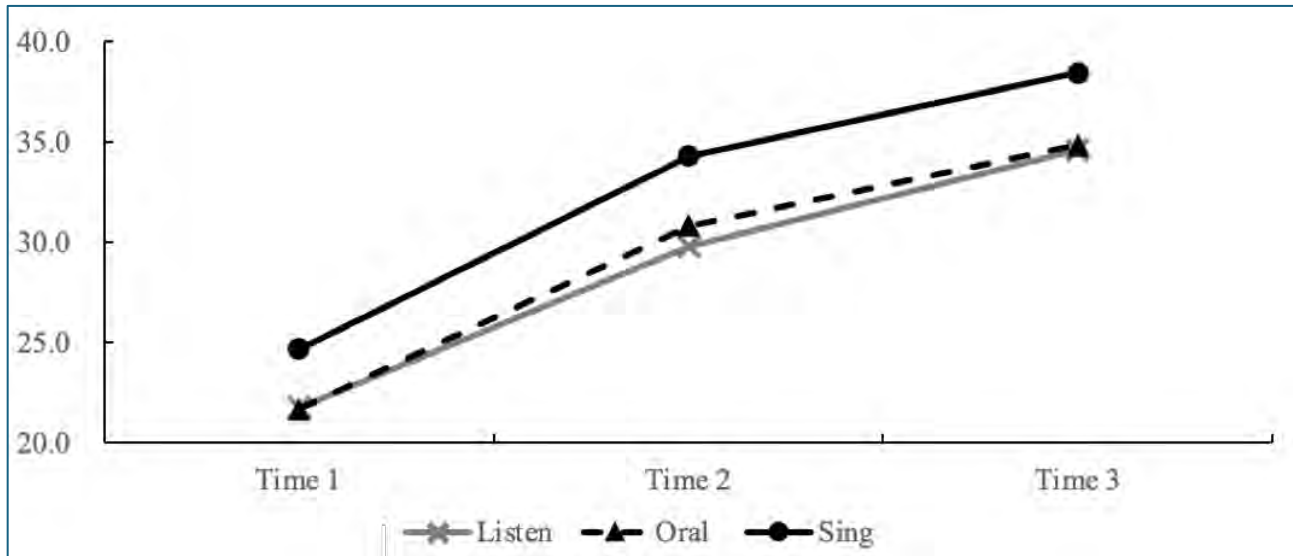**Figure 2. The Interaction Effects Between Vocabulary Dimensions and Times**

Figure 3. The Interaction Effects Between Different Songs and Times

Table 6. Summary of the Analysis of HLM

Fixed effects	B (95%CI)	SE	t	p
Intercept	39.71 (34.50 – 44.93)	2.66	14.93***	<.001
Time				
2 vs 1	5.27 (-0.17 – 10.70)	2.77	1.90	.058
3 vs 1	12.43 (7.59 – 17.27)	2.47	5.03***	<.001
Practice modes				
O vs L	-0.08 (-1.87 – 1.71)	0.91	-0.09	.930
S vs L	2.92 (0.81 – 5.02)	1.07	2.73**	.007
S vs O[a]	3.00 (0.84 – 5.16)	1.10	2.73**	.007
Dimensions				
WMR vs AMR	6.19 (4.26 – 8.13)	0.99	6.27***	<.001
WFP vs AMR	-21.44 (-24.36 – -18.51)	1.49	-14.37***	<.001
WFP vs WMR[a]	-27.63 (-30.95 – -24.31)	1.69	-16.33***	<.001
Songs				
2 vs 1	-5.59 (-7.75 – -3.43)	1.10	-5.08***	<.001
3 vs 1	-9.74 (-11.76 – -7.73)	1.03	-9.47***	<.001
3 vs 2[a]	-4.15 (-6.04 – -2.26)	0.96	-4.30***	<.001
Time * Practice modes				
O vs L (T2-T1)	1.08 (-1.29 – 3.45)	1.21	0.90	.370
S vs L (T2-T1)	1.58 (-0.67 – 3.83)	1.15	1.37	.169
S vs O (T2-T1) [a]	0.50 (-1.74 – 2.73)	1.14	0.43	.664
O vs L (T3-T1)	0.33 (-1.93 – 2.58)	1.15	0.28	.777
S vs L (T3-T1)	0.94 (-1.22 – 3.11)	1.10	0.85	.393
S vs O (T3-T1) [a]	0.62 (-1.57 – 2.81)	1.12	0.55	.580
Time * Dimensions				
WMR vs AMR (T2-T1)	2.32 (-1.90 – 6.54)	2.15	1.08	.281
WFP vs AMR (T2-T1)	1.43 (-2.30 – 5.15)	1.90	0.75	.453
WFP vs WMR (T2-T1) [a]	-0.90 (-5.48 – 3.69)	2.34	-0.38	.702
WMR vs AMR (T3-T1)	1.10 (-1.89 – 4.09)	1.52	0.72	.471
WFP vs AMR (T3-T1)	-4.75 (-9.80 – 0.30)	2.57	-1.85	.065
WFP vs WMR (T3-T1) [a]	-5.85 (-11.59 – -0.11)	2.93	-2.00*	.046
Time * Songs				
2 vs 1 (T2-T1)	1.81 (-0.22 – 3.85)	1.04	1.75	.080
3 vs 1 (T2-T1)	2.75 (0.36 – 5.14)	1.22	2.26*	.024
3 vs 2 (T2-T1) [a]	0.94 (-1.48 – 3.36)	1.23	0.76	.447
2 vs 1 (T3-T1)	2.16 (0.00 – 4.33)	1.11	1.96	.051
3 vs 1 (T3-T1)	2.61 (0.31 – 4.92)	1.17	2.22*	.026
3 vs 2 (T3-T1) [a]	0.45 (-1.69 – 2.58)	1.09	0.41	.681
Covariate				
2 vs 1	-9.55 (-15.77 – -3.32)	3.17	-3.01**	.003
3 vs 1	-13.68 (-19.62 – -7.74)	3.03	-4.52***	<.001

Random Effects	Variance	SE	Z	p
Level II (error, between group)	124.892	20.9950	5.961***	<.001
Level I (error, within group)	293.38	8.765	33.474***	<.001

Note: [a] Reference group changed for the second HLM. WMR: written meaning recognition, AMR: aural meaning recognition, WFP: written form production
*p < .05, **p < .01, ***p < .001

Summary of Interview Results

The interview focused on four main questions, and students' reflections and the frequencies for each question are briefly summarized as below. Students A-G were female, and H-L were male.

How did you like the three songs we practiced in class? Only students D and I expressed that they liked the melody of "The Sound of Silence." Seven students did not particularly show any fondness for the selected songs, and three students disliked all the songs. Two major reasons for not having particular fondness for these songs were that the tempos of these songs were slow and dull, and the lyrics contained many unfamiliar words, such as tenement, sycamore, and daffodil. All the interviewees reported they had never heard the three songs before, and the concepts of these songs seemed abstract and difficult for them to interpret.

Did you continue practicing the three songs after the intervention? All students expressed the song that they remembered best was the song they had to sing, and six of the students reported that they occasionally sang a few lines that they could remember. Student J commented, "It is very difficult to remember all the lyrics even after I had practiced the song many times; "Starry, Starry Night" was particularly difficult." Two higher-level students reported that it was difficult for them to orally read the lyrics after listening to them, and none of them read the lyrics after the intervention was over. Student F reflected "songs are to be sung, not orally read, so I did not read them orally; I just listened to them a few times when I was on the bus." Student L expressed that he liked watching Venessa Williams singing "The Colors of the Wind," but it was difficult for him to sing it. Overall, some students sang only parts of the songs that they had learned to sing, and none orally practiced the song lyrics.

Did you find it very anxiety-provoking to record your singing or oral reading and present it to your teacher? All the student interviewees reflected that singing was difficult and embarrassing for them, and none considered themselves good at singing. Two students, however, expressed that they thought they could have sung better if they had not been required to video-record their singing. Three students expressed that they liked listening to songs only, but not singing them or orally reading the lyrics; the main reason, according to them, was that they did not have an ear for music, let alone the ability to sing well enough to do justice to the songs. Seven students reflected that even after they had listened to songs, they were still not sure about the pronunciations of many words because words sounded different in singing and in speaking. All these reasons made the singing and oral reading anxiety-provoking.

What did you think you learned from these songs? All students reported that they had learned many words during the intervention period, but most words were difficult to remember. Only student D reported that listening to these songs improved her listening skills, and she also learned some phrases. Student I reflected that he preferred to learn English through playing video games than by listening to music.

To summarize students' reflections about the intervention, none of the students were familiar with the three songs and they did not particularly favor any of the songs; from time to time, they sang a few lines of the songs they could remember, but not the whole song. Requiring students to video-record their singing or oral reading of the lyrics was an anxiety-provoking task. Most students considered that they learned only vocabulary words and nothing else from these songs.

Discussion

The previous section has shown that the four fixed variables have significant main effects on L2 learners' retention of vocabulary from songs. In this section, the effects of each variable will be discussed.

The Time Effect on the Retention of Vocabulary from Songs

The results of the study showed that after the varying practice modes for six weeks, students gained an average of 9.92% (5.95 words = 9.92*.6) for studying three songs; their scores improved from 21.91% at Time 1 to 31.83% at Time 2; however, the gain from Time 1 to Time 2 was statistically nonsignificant. Interestingly, students gained another 4.29% from Time 2 to Time 3, making a total of 14.21% (8.53 words) from Time 1 to Time 3, and the difference between the two periods was statistically significant. Although it might not be surprising that students performed better on the 8-week delayed post-test than on the 1-week delayed post-test due to practice effect (see student interviewees' reflections below and Pavia et al., 2019, Webb & Chang, 2022), after-test discussion (Nation, 2013) or searching for answers, this study, however, found another reason for the further gain after the end of the treatment. That was, according to the interviewees, six of the interviewees reporting that they occasionally sang or listened to the target songs they learned even after the end of intervention.

The total advancement from Time 1 to Time 3 from learning three songs was not quite as satisfactory as those reported by Baills et al. (2021) and Pavia. (2023) due to many differences in the research design between the present study and those of previous ones. One salient difference was that in the present study, songs were integrated in their normal listening course; students were not informed that they would be tested, nor were they awarded anything for doing the tasks. The present study, therefore, is more ecologically valid as compared to the one by Baills et al. (2021), whose study was more like a lab-controlled study and did not look at the longer-term effect. Another difference was that the three songs were far more difficult than the song used in Baills et al. in terms of vocabulary levels. In contrast to Pavia's study, only one song was used in each of her three sub-studies. The present study was therefore more demanding in terms of learning load. Although this could be one of the drawbacks of the present study if the learning rates of individual songs were compared, if all new words acquired were tallied up, the results could be quite encouraging. Finally, the songs used in the present study were not selected by students themselves, which might have demotivated them from doing their best in acquiring vocabulary.

The Different Practice Modes of Song on Vocabulary Learning

As shown in Table 6, students gained significantly more from the singing than from oral reading and listening. The results were different from those of the study by Davis and Fan (2016), who found no differences among the three modes (sings, choral repetition, or neither) for young children but corroborated Ludke et al. (2014), who found the effect of singing was superior to the rhythmic speaking and speaking conditions. Among the three practice modes, singing is supposed to require higher level of involvement load than listening only. According to student interviewees, all students reflected that singing was very difficult for them, especially for the low-level learners and those who are not musically inclined. They reported that they had to learn the meaning and pronunciation of the unknown words as well as the melody of the songs at the same time. The cognitive load of singing was apparently very heavy for the low-level students. Despite the negative comments, the effort paid off because the singing mode led to higher gains. The researcher must admit that the three selected songs were somewhat difficult for the lower-level students and

this might be the main reason that the retention rate of students in the study was lower than those of Baills et al. (2021), whose students learned only 14 words for body parts.

Although both singing and oral reading modes involved one extra round of engagement, the vocabulary gain did not differ significantly between the oral reading and listening modes. According to students' report, asking students to orally read the lyrics was a bit awkward after they had listened to the songs. Two higher-level student interviewees reflected that they felt it difficult to orally read the lyrics after they had listened to the songs a few times because they would rather sing than read aloud the parts that they could sing. From students' report, it might be a better idea to ask students to become familiar with the lyrics through reading orally, which means that students may orally practice the pronunciation of the unknown words first then practice the sentences until they can orally read the lyrics fluently. Afterwards, students can listen to the song, and finally sing along. This strategy seems to break down the learners' cognitive load into three stages. Whether following the order of input will improve students' learning is unclear and may have to rely on future studies to explore.

Finally, Table 5 shows students gained slightly less in the listening mode than the oral reading mode from Time 1 to Time 2, and from Time 1 to Time 3; however, attention should be paid to the gain from Time 2 to Time 3. Students gained slightly more through listening than through oral reading and singing, with 4.77%, 4.01%, and 4.13% for listening, oral reading and singing modes respectively. Although listening required less cognitive load than singing and oral reading; all student interviewees gave no negative comments about listening to the selected songs. Listening only is supposed to be very relaxing and enjoyable, and its effect might be gradual.

The Retention of Different Vocabulary Dimensions through Songs

The students consistently performed best on written meaning recognition (WMR) across the time points, followed by aural meaning recognition (AMR) and written form production (WFP), and the results were consistent with previous studies in that recognition tasks are always easier than production tasks, in particular for word meaning (Coyle & Gracia, 2014; Laufer & Rozovski-Roitblatv, 2015; Teng, 2019). As well, many word meanings had been taught by the instructor before the intervention, and every week the students had an opportunity to read the lyrics while listening, reading orally, or singing along. These repetitions might have helped students to consolidate the word meanings more easily than the word forms. Students scored the lowest in WFP. The main reason could be that the task was too demanding because while listening to the sentences, students had to choose a correct meaning from one of the six options (not four options as in most of the studies), and then write down the spelling of the word they heard. Their attention was split into three parts: listening for the sentence meaning, reading and choosing a correct L2 meaning, and producing the correct written form. All the tasks had to be completed within10 seconds otherwise they missed out on the next item. Since recognizing an L2 meaning is easier than producing the correct form. Many students might have chosen the easier task and left the more difficult one undone or uncompleted. Finally, this study adopted a strict scoring approach; that is any mistake, e.g., spelling dafodil for daffodil, resulted in zero points. The above reasons explained why students scored lowest on WFP throughout. The result was corroborated Coyle and Gracia (2014), who found that giving three 30-minute song exposures did not help young children's productive knowledge.

Another observation that worth mentioning is from Time 2 to Time 3, students continued to make advancement with 7.68%, 5.55%, and 0.59% for AMR, WMR, and WFP respectively. The higher

gain for the dimension of AMR after the intervention could most likely due to students relistening to the songs because previous studies have shown that the more repetitions, the higher rate of acquisition (see Uchihara et al., 2019 for a review). When students did more listening to these songs after the intervention, that might have helped them better comprehend the lyrics of the songs, especially on the meanings of the lyrics. Increasing the repetitions of listening might facilitate L2 learners' knowledge, in particular, on recognizing meanings. Despite students making more advancement after the intervention, the overall retention rate is less satisfactory than learning vocabulary from reading graded readers or listening to stories (see, Webb & Chang, 2022). From students' scores, researcher's observations and interviewee's reflections, learning vocabulary through songs may take time to see its effect grow.

The Effect of Varying Songs on Vocabulary Retention

The three songs selected in the present study contained many unfamiliar words, and 28/60 target words were beyond 3,000-word level. In the pre-test, students scored lowest for Song 3, "The Colors of the Wind," (18.09%) and the highest for Song 1, "The Sound of Silence," (27.83%), but the learning rates for Song 3 were significantly higher than for Song 1 from the pre-test to the 1-week and 8-week delayed post-tests. One main reason for the significant difference in gains could be that many unknown words from Song 3 are short and concrete nouns, such as otter, heron, hoop, trail whereas words from Song 1 are long or abstract, e.g., prophet, halo, tenement, vision. This phenomenon seemed to corroborate previous studies that showed that abstract nouns are more difficult to learn than concrete nouns (Chang, 2019; De Groot & Keijzer, 2000; Ellis & Beaton, 1993; van Hell & Candia Mahn, 1997). Compared to the study by Pavia et al. (2019), their students in the experimental groups also performed differently for different songs. One of their explanations was that the target words in the two songs were different in their levels of difficulty. This reason seems also applicable to the present study.

In addition to the difference in lexical difficulty, another reason reported by the student interviewees was that they preferred a song with a faster rhythm; they considered songs with slow tempo "dull." Song 1 was sung at the lowest speech rate, 58 words per minute (WPM), Song 2, 76 WPM, and 81 WPM for Song 3, so students' preferences seem to have played some role in their learning. Other factors, e.g., repetition and performers, are not within the scope in this study.

Conclusion

This study provided some more hard evidence into the retention of English vocabulary through songs. The findings, which provided answers to the two research questions, are summarized as follows:

1. All the four fixed variables (time, practice mode, vocabulary dimensions, and varying songs) were found to have significant effects on the retention of vocabulary. Students made significant advancement from Time 1 to Time 3, but not from Time 1 to Time 2. Singing mode was found more effective than listening and oral modes. Students also achieved significantly better on WMR than on AMR and WFP.

2. The interaction effects between time and input mode, and between time and vocabulary dimensions were nonsignificant. The interaction effects between time and songs were found significant between Song 1 and Song 3 from Time 1 to Time 2, and from Time 1 to Time 3, which may imply that some words in certain songs may be easier to acquire and to retain, or that preferred songs invite a larger number of repeated listening.

On the whole, students did retain some vocabulary words through English songs, but the rates were not quite as satisfactory as learning vocabulary through reading or listening materials. It is certain that the total time for the input was not sufficient (approximately 60 minutes per song), but this study found that songs had a long-lasting effect and students were more willing to listen repeatedly than to sing or to orally read the lyrics. It is likely that there are other benefits, such as facilitating L2 lexical stress processing (Degrave, 2022), pronunciation (Moradi & Shahrokhi, 2014), etc., but they are not within the scope of this study. All these advantages may imply that songs can be integrated into the course, but it takes time to see its effect growing.

Before ending this paper, some limitations of the paper should be pointed out. Firstly, all students practiced the same three songs; this might not be fair to the lower-level students because there are many aspects that they have to learn at the same time, which include learning meaning, pronunciation, spelling, song beats, etc. To improve the research design, future studies may include a few songs with different levels of difficulty for students to select or include language proficiency to examine the role of language proficiency on the effects of learning a language through songs. Secondly, this study involved three practice modes but all were determined by draws. This course of action did not consider individual differences. In real-life situations, learners can determine their favorite way to practice songs; therefore, future studies may improve the research design by allowing students to choose their preferred way of practice. Thirdly, the selected songs in the study were determined by English teachers and some senior students, which might have overlooked students' musical preferences. This limitation might have affected students' motivation to learn the songs. Future studies may conduct a survey of the students of the same age to explore their favorite songs and performers. Lastly, to asking them to upload their singing and to send their songs to their teachers is very anxiety-provoking for many students. A better way, if singing is necessary, is to sing in groups instead of individually. Since the effect of learning linguistic elements from songs is not easily determined in many short-term studies, a long-term study is warranted.

About the Author

Anna C-S Chang is a Professor in the Department of Applied English at Hsing Wu University, New Taipei, Taiwan, teaching courses on English listening, reading and vocabulary. Her main research interests focus on listening and reading development, and vocabulary learning. She has published extensively with internationally refereed journals. ORCID ID: 0000-0003-4216-6394

Acknowledgements

To Cite this Article

Chang, C-S. (2024). The effects of varying practice modes on L2 learners' vocabulary retention through songs: listening, singing, and oral reading. *Teaching English as a Second Language Electronic Journal (TESL-EJ), 28*(3). https://doi.org/10.55593/ej.28111a9

References

Baills, F., Zhang, Y., Cheng, Y., Bu, Y., & Prieto, P. (2021). Listening to songs and singing benefitted initial stages of second language pronunciation but not recall of word meaning. *Language Learning*, *71*(2), 369-413. https://doi.org/10.1111/lang.12442

Chang, C-S (2019). Effects of narrow reading and listening on L2 vocabulary learning: Multiple dimensions. *SSLA* 41(4), 769-794. https://doi.org/10.1017/S0272263119000032

Chou, M. H. (2014). Assessing English vocabulary and enhancing young English as a Foreign Language (EFL) learners' motivation through games, songs, and stories. *Education 3-13*, *42*(3), 284-297. https://doi.org/10.1080/03004279.2012.680899

Coyle, Y., & Gómez Gracia, R. (2014). Using songs to enhance L2 vocabulary acquisition in preschool children. *ELT Journal*, *68*(3), 276-285. https://doi.org/10.1093/elt/ccu015

Davis, G. M., & Fan, W. (2016). English vocabulary acquisition through songs in Chinese kindergarten students. *Chinese Journal of Applied Linguistics*, *39*(1), 59–71. https://doi.org/10.1515/cjal-2016-0004

Degrave, P. (2022). Music training and the use of songs or rhythm: Do they help for lexical stress processing? *International Review of Applied Linguistics in Language Teaching, 60*(3), 799-824 https://doi.org/10.1515/iral-2019-0081

De Groot, M. B., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of concreteness, cognate, status. and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning*, *50*(1), 1-56. https://doi.org/10.1111/0023-8333.00110

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, *43*(4), 559-617. https://doi.org/10.1111/j.1467-1770.1993.tb00627.x

Hirata, M. C. (2018). Potential for incidental vocabulary learning in songs: A Corpus-based study. *Journal of Foreign Language Education and Technology*, *3*(2), 2536-4731.

Huta, V. (2014). When to use hierarchical linear modeling. *The quantitative methods for psychology, 10*(1), 13-28. https://doi.org/10.20982/tqmp.10.1.p013

Hyman Jr, I. E., Burland, N. K., Duskin, H. M., Cook, M. C., Roy, C. M., McGrath, J. C., & Roundhill, R. F. (2013). Going Gaga: Investigating, creating, and manipulating the song stuck in my head. *Applied Cognitive Psychology*, *27*(2), 204-215. https://doi.org/10.1002/acp.2897

Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research*, *19*(6), 687-711. https://doi.org/10.1177/1362168814559797

Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, *10*(1), 105-129. https://doi.org/10.1080/14790718.2012.679275

Ludke, K. M., Ferreira, F., & Overy, K. (2014). Singing can facilitate foreign language learning. *Memory & cognition*, *42*(1), 41-52. https://doi.org/10.3758/s13421-013-0342-5

Moradi, F., & Shahrokhi, M. (2014). The effect of listening to music on Iranian children's segmental and suprasegmental pronunciation. *English Language Teaching*, *7*(6), 128-142. https://doi.org/10.5539/elt.v7n6p128

Murphey, T. (1992). The discourse of pop songs. *TESOL Quarterly*, *26*(4), 770-774. https://doi.org/10.2307/3586887

Nation, I. S. P. (2013). Learning vocabulary in another language (2nd ed.). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139858656.002

Pavia, N., Webb, S., & Faez, F. (2019). Incidental vocabulary learning through listening to songs. *Studies in Second Language Acquisition*, *41*(4), 745-768. https://doi.org/10.1017/S0272263119000020

Pavia, N. (2023). *The Effects of Song use on Vocabulary Learning: Studies of Distribution of Practice, Modes of Input, Retrieval, and the Input-Output-Input Sequence of Exposure* (Doctoral dissertation, The University of Western Ontario (Canada).

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition*, *106*(2), 975-983. https://doi.org/10.1016/j.cognition.2007.03.005

Tegge, F. (2017). The lexical coverage of popular songs in English language teaching. *System*, *67*, 87-98. https://doi.org/10.1016/j.system.2017.04.016

Tegge, F. (2018). Pop songs in the classroom: Time-filler or teaching tool? *ELT Journal*, *72*(3), 274–284. https://doi.org/10.1093/elt/ccx071

Teng, F. (2019). The effects of video caption types and advance organizers on incidental L2 collocation learning. *Computers & Education,* 142. https://doi.org/10.1016/j.compedu.2019.103655

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language learning*, *69*(3), 559-599. https://doi.org/10.1111/lang.12343

van Hell, J. G., & Candia Mahn, A. (1997). Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and inexperienced learners. *Language Learning*, *47*(3), 507-546. https://doi.org/10.1111/0023-8333.00018

Webb, S., & Chang, C-S (2022). How does mode of input affect the incidental learning of collocations? *Studies in Second Language Acquisition*, 44(1), 35-56. https://doi.org/10.1017/S0272263120000297