Language Learning & Technology
ISSN 1094-3501 CC BY-NC-ND

2022, Volume 26, Issue 1
pp. 1–24

ARTICLE

Association between the characteristics of out-of-class technology-mediated language experience and L2 vocabulary knowledge

Chun Lai, The University of Hong Kong

Yang Liu, Tianjin University

Jingjing Hu, Sun-Yat Sen University

Phil Benson, Macquarie University

Boning Lyu, Xiamen University

Abstract

Out-of-class autonomous language learning with technology is positively associated with learners' L2 vocabulary knowledge (Lee, 2019; Webb, 2015). An understanding of how out-of-class technology-mediated language experiences relate to L2 vocabulary development is essential to discussions about the quality of out-of-class language learning experiences. This study examined 46 Chinese EFL learners' self-directed out-of-class language learning experiences with technology in order to develop a framework of the defining characteristics of out-of-class technological experiences that are associated with L2 English vocabulary knowledge. Analysis of the learners' one-month-long diaries recording their daily technology activities in English, semi-structured interviews, and performance in a vocabulary knowledge assessment revealed several characteristic indicators that were positively associated with L2 English vocabulary scores. It was found that accessing multimodal materials, dual attention to meaning and form, the depth of lexical information attended to and the levels of engagement with words when engaging in technology activities were significantly associated with L2 English vocabulary scores. The findings suggest these dimensions as potential directions for future research and as core aspects of learner support for out-of-class language learning.

Keywords: *Out-of-class Language Learning, Vocabulary Learning, Autonomous Language Learning*

Language(s) Learned in This Study: *English*

APA Citation: Lai, C., Liu, Y., Hu, J., Benson, P. & Lyu, B. (2022). Association between the characteristics of out-of-class technology-mediated language experience and L2 vocabulary knowledge. *Language Learning & Technology*, *26*(1), 1–24. https://hdl.handle.net/10125/73485

Introduction

Autonomous language learning beyond the classroom is common among language learners (De Wilde et al., 2020; Toffoli & Sockett, 2010). Technology provides the learning spaces and tools to facilitate learners' out-of-class engagement with English (Lai, 2017; Steel & Levy, 2013). Previous research has shown that language learners voluntarily engage in a wide variety of technology activities beyond the classroom, and positive relationships have been identified between these activities and various aspects of language proficiency, including L2 vocabulary knowledge (Cole & Vanderplank, 2016; De Wilde et al., 2020; Peters, 2018). However, different technological experiences hold differential potential for language learning (Lai et al., 2018; Sundqvist & Sylvén, 2016). For instance, TV shows are found to afford greater contextual cues for language learning than songs or news (Toffoli & Sockett, 2010). Different registers of the same technological resource also provide different affordances for language learning. For instance, sitcoms are found to be closer to natural conversation than soap operas (Al-Surmi, 2012). Even differential use of the same type of technological resource may lead to different learning outcomes. For instance, narrow viewing (e.g., watching different episodes of the same program) is shown to afford more opportunities for vocabulary learning than watching unrelated programs (Rodgers & Webb, 2011). Thus, the extent to which out-of-class technological experiences benefit language development depends fundamentally on how learners construct their experiences. Insights into how different types of technological experiences relate to language learning inform the conceptualization of the quality of out-of-class language learning experiences, an understanding that is essential to guiding learners' construction of beneficial language learning experience beyond the classroom.

To shed light on this research issue, the present study examines a group of Chinese EFL university students' autonomous out-of-class technology-mediated language learning activities to reveal some characteristics of the activities that are positively associated with vocabulary knowledge. The study aims to enrich the current literature on autonomous out-of-class language learning with technology in two ways. First, it extends the research discourse beyond providing evidence of the nature of learners' experiences and the benefits thereof towards a greater focus on the quality of the experiences (Blyth & LaCroix-Dalluhn, 2011; Lai et al., 2018). Second, it provides a holistic yet in-depth understanding of the characteristics of out-of-class technological experiences and their relation to L2 vocabulary development, a research issue that is not well explored but the answer to which is essential to "maximize[ing] the benefits of this exposure to vocabulary acquisition" (Schmitt, 2019, p. 9).

Literature Review

Out-of-Class Language Exposure and Vocabulary Acquisition

Extending vocabulary learning beyond the classroom is crucial to developing the large amount of vocabulary learners need in order to comprehend everyday audiovisual sources and written texts (Nation, 2015; Schmitt, 2014; Webb & Nation, 2017). Exposure to language outside the classroom also creates essential conditions for vocabulary development–diverse environments and abundant opportunities for incidental learning (i.e., vocabulary learning as a by-product of meaning-focused activities without deliberate intention to learn) (Godwin-Jones, 2018; Webb & Nation, 2017). Out-of-class exposure to the L2 has been found to be a more significant predictor of vocabulary knowledge than L2 instruction and L2 immersion (Garnier & Schmitt, 2016; Peters, 2018).

Technology plays a significant role in learners' out-of-class engagement with the language. Empirical studies have attested to the positive relationships between out-of-class technology activities and vocabulary knowledge (Lee, 2019; Webb, 2015). Different technological resources have differential potential for vocabulary learning (De Wilde et al., 2020; Peters, 2018). For instance, Peters (2018) reported that vocabulary knowledge had positive correlations with the use of un-subtitled movies and TV programs, books, magazines, and websites; a negligible negative correlation with songs; and no correlations with subtitled movies and TV programs and computer games. De Wilde and colleagues (2020) further found that activities involving the active use of English for communication (e.g., the use of English language social media and oral communication in English) were significant predictors of receptive vocabulary knowledge, whereas less interactive activities, such as listening to music and reading, did not significantly predict children's vocabulary knowledge. In addition, in Sundqvist's (2019) study, different ways of interacting with the same technological resource (e.g., multiplayer gaming vs. single player gaming) were found to result in different gains in vocabulary. The findings from this body of literature were rather mixed. For instance, some studies reported that watching TV led to greater gains in vocabulary knowledge than playing computer games (e.g., Lindgren & Muñoz, 2013), while other studies found that gaming predicted students' receptive vocabulary knowledge but watching TV did not (e.g., De Wilde et al., 2020). Inconsistent findings were also reported in studies that examined the effects of modality on vocabulary learning. For instance, Vidal (2011) found that reading academic texts resulted in higher levels of incidental vocabulary learning than listening to academic texts. Arndt and Woore (2018) reported contrary findings in that watching video blogs outperformed reading blog posts in promoting both knowledge of the grammatical functions of words and meaning recognition and recall. Feng and Webb (2020) found no significant differences in incidental vocabulary learning through viewing TV show episodes, listening to the audio narration of the episodes and reading the written transcripts of the episodes.

The inconsistent findings suggest that using individual technological resources as the units of analysis might be problematic since the same technological resource might be interpreted and used differently by learners with different learning beliefs and for different learning purposes (Ala-Kyyny, 2012; Lai, 2015, 2019; Vanderplank, 2019). Moreover, learners' perceptions of their informal technological experiences may shape their agency and efforts in language learning when interacting with the resources (Lyrigkou, 2019). Thus, a learner-driven approach that features learner-defined technological use based on learner needs and interests might provide more nuanced insights into informal digital media as learning resources (Arndt & Woore, 2018; Vanderplank, 2019). A systematic framework with heuristics that categorizes the defining properties of out-of-class technological experiences associated with vocabulary knowledge might be more informative, and generate a nuanced understanding of the relationship between technological experiences and vocabulary development.

Frameworks of Out-of-Class Language Learning with Technology

Benson (2011) and Richards (2015) were among the first scholars to propose categorizing frameworks to differentiate out-of-class learning activities. Benson (2011) categorized out-of-class language learning along four dimensions: (a) the degree to which the learning experience is educationally structured; (b) the degree to which learners perceive learning as self-controlled; (c) the physical, social, and pedagogical relationships between learners and the learning environment; (d) and the extent to which instructional processes are involved. Richards (2015) added a few more dimensions pertaining to the nature of the communication medium and the characteristics of the task and language use. However, both frameworks focus on out-of-class learning in general and are inadequate to capture the nuanced differences in self-directed technological activities. Also, both are based on analyses of activities and fail to consider learners' agentic reinterpretation of, and engagement in, the activities. Lai and colleagues (2018) constructed a framework specifically for self-directed out-of-class language learning with technology, categorizing technological activities according to learner-defined interaction purposes and interaction patterns with the resources. The framework features three distinct types of technology activities including instruction-oriented, entertainment- and information-oriented, and socialization-oriented experience. Despite being situated in the context of autonomous out-of-class technology-mediated language learning, this framework, similar to the other two existing frameworks, is descriptive in nature without relating the categorization to specific language learning outcomes. Relating the nature of technology activity to specific language learning outcomes is important because the optimal conditions for the development of different aspects of language may vary. A framework concerning the quality of out-of-class technological use makes more sense when it relates to specific language outcomes.

To address the limitations of existent frameworks, this study adopted a learner-centric approach to uncover some underlying properties of out-of-class technological activities associated with vocabulary knowledge. Such a framework acts as a guide to the construction of effective vocabulary learning experiences beyond the classroom. We referred to the literature on essential conditions for vocabulary learning in conceptualizing the potential properties.

Theoretical Framework

Vocabulary learning is "mainly a matter of selective attention and elaborated processing" (Hulstijn, 2013, p. 2634). Noticing or conscious registration of words is necessary for 'input' to turn into 'intake,' and whether and how the noticed intake is further processed (i.e., the depth of processing) determine the learning process (Leow, 2018). Thus, more noticing and deeper processing lead to more learning (Hulstijn, 2013; Schmidt, 2010). In Leow's (2018) words, "the higher the level of awareness and the greater the depth of processing, the more likely it is that robust learning will take place" (p. 3). The two essential and complementary types of vocabulary learning—incidental vocabulary learning and intentional vocabulary learning (i.e., vocabulary learning through the utilization of retention techniques to commit lexical information to memory)—both involve attention and noticing, but differ in the aspects of formal and semantic features of words that are attended to, and in how deeply they are processed (Webb, 2020). Incidental and intentional vocabulary learning with different language sources and activities, including different technology activities, activates the noticing and processing of different lexical information (Sundqvist & Sylvén, 2016; Webb, 2020). Thus, the frequency of incidental exposure to the lexical items via varied resources and experiences, the degree of deliberate learning of words, and the attention paid to and the cognitive operations performed on words are highlighted in seminal discussions about vocabulary learning (Laufer, 2017; Nation, 2015; Schmitt, 2008). In essence, two major dimensions are underscored in quality vocabulary learning experience: (a) balanced resources and activities of a diversified nature; and (b) depth of engagement with vocabulary. These two dimensions served as the theoretical lens to identify the characteristics of quality autonomous out-of-class technology activities in this study.

For the balance and diversification dimension, the modality of the technological resource (i.e., audio, visual, and textual) might be a relevant factor since resources of different modalities have differential vocabulary learning potential, and resources that contain audiovisual plus textual information may benefit vocabulary learning more (Arndt & Woore, 2018; Feng & Webb, 2020). The opportunity for both inductive learning of vocabulary in context and decontextualized explicit vocabulary study might be another relevant factor (Godwin-Jones, 2018; Toffoli & Sockett, 2010). The balance of meaning and form-focused activities might be yet another essential factor (Lee, 2019) because meaning-focused activities are more conducive to the contextual aspects of word knowledge, while form-focused activities are beneficial to the most salient aspects of word knowledge, such as the form-meaning link (Godwin-Jones, 2018; Laufer, 2017; Schmitt, 2014). For the depth of engagement dimension, the extent to which the technological activity induces attention to different types of lexical information (the form-meaning link or lexical association and usage) matters (Schmitt, 2014; Sundqvist, 2011). Moreover, according to Laufer and Hulstijn's (2001) Involvement-Load Hypothesis, the levels of learners' motivational involvement (i.e., the 'need' to engage with the word) and cognitive involvement (i.e., 'searching' for the meaning, form or usage of a word, and making an 'evaluative' judgement on the appropriateness of the meaning, form or usage of a word) in processing a word determine the likelihood of the word being acquired and retained. Levels of engagement with vocabulary hence might be another essential factor in this dimension. Thus, the current literature suggests five indictors of quality language experience with regard to vocabulary development: three indictors along the balance and diversification dimension (multimodality, the balance of contextualized exposure and decontextualized study, and balanced attentional focus); and two indictors along the depth of engagement dimension (depth of lexical information, and involvement load).

Guided by the theoretical framework, this study applied the five indicators to analyzing the out-of-class technology-mediated English language activities of a group of university undergraduate students. It examined how the five indicators of technological activities were associated with vocabulary knowledge. Specifically, the study addressed the following two research questions:

> RQ1: What characteristic indicators of the balance and diversification dimension were associated with participants' vocabulary scores?
>
> RQ2: What characteristic indicators of the depth of engagement dimension were associated with participants' vocabulary scores?

Method

Participants

Forty-seven undergraduate students (15 males and 32 females) studying a variety of majors (e.g., engineering, education, law, chemistry, management, medicine) were recruited through their English instructors from two comprehensive research universities in mainland China, one in the northern interior region and the other in the eastern coastal region. The coastal regions in China are economically more developed and integrated into the global economy than the interior regions, have greater societal expectations concerning the English language, and enjoy more English language educational resources (Clothey, 2012). Recruiting participants from universities in different regions allowed for greater variation in vocabulary knowledge and extramural English experience to facilitate the research inquiry and extend the ecological validity of the study. Because a precondition for the study was students' normal autonomous out-of-class engagement with the English language, the recruitment targeted students who were not planning on attending English public tests in the next few months and who self-reported engaging in regular technology-mediated English language activities beyond the classroom. Forty-two participants were second-year undergraduate students taking English for academic purposes classes, two were first-year students taking general English classes, and three were in their third year of undergraduate study and were not taking formal English classes. The participants had an average of 8 years of prior English learning experience. All but two of the participants had passed the CET-4 (College English Test Band 4), a national English test, which is equivalent to CEFR (Common European Framework Reference) B2 level, and half of the participants had passed the CET-6, which is equivalent to CEFR C1 level (Hou, 2017). Passing the CET-6 English test is a common societal expectation for university students in China, and some of the participants also had plans to take the TOEFL or IELTS test in the future.

Data Collection and Procedure

The study lasted six weeks and involved three data sources: (a) a vocabulary test, (b) a diary of daily technology-mediated English language activities, and (c) a semi-structured interview.

Vocabulary Test

The participants, upon agreeing to participate in the study, took a vocabulary test. The 62-item test battery measured both receptive and productive vocabulary size (32 items) and vocabulary depth (30 items). The dimensions and sources of the test items are summarized in Table 1. Items were taken from several test batteries rather than adopting complete validated tests because we wanted to include multiple vocabulary breadth and vocabulary depth measures (Schmitt, 2014), but at the same time make the test length manageable. The vocabulary items were selected in such a way as to ensure the tested vocabularies were from a wide frequency range, well beyond Chinese university students' estimated average vocabulary size of around 4200 words (Hui, 2004), and contained different word classes. The test was administered under the supervision of one of the researchers. The participants were given 30 minutes to complete the test, and all participants completed the vocabulary test within the given time with most of them finishing within 25 minutes.

Table 1

The Dimensions and Sources of the Test Items

	Sources of Test	Type of Question	Test Items
Vocabulary Size Tests	Vocabulary Size Test (Nation, 2012)	10 multiple-choice items	2 items from each of the 3rd to the 7th 1000 word levels
	Vocabulary Levels Test (Nation, 1990)	12 multiple-choice items	3 items from each of the 2000, 3000, 5000 and 10,000 word levels
	PVLT test (Laufer & Nation, 1999)	10 fill-in-the-missing-letter items	2 items from each of the 2000 and 2000-3000 word levels; and 3 items from each of the 5000 word level and the university word list
Vocabulary Depth Tests	Polysemous Phrasal Verb Test (Garnier & Schmitt, 2016)	10 fill-in-the-missing-letter items	150 most frequent phrasal verbs with a wide range of the frequencies of the meaning senses: 3 items below 1000; 5 items between 1000-5000; and 2 items above 10000
	Derivatives Word Form Test (Schmitt & Zimerman, 2002)	10 fill-in-the-blank items	all items from the academic vocabulary list
	Word Association Test (Read, 1998)	10 multiple-choice items	3 items from the 1000 word level; 2 items from the 2000 word level; and 5 items from the academic vocabulary list

Diary

The participants were then asked to keep an electronic diary, in a word document, recording their normal autonomous out-of-class technology-mediated English activities for one month. They were also asked to take screenshots or video recordings of representative technology resources/activities that they felt influenced their vocabulary development. To enhance the richness of the diary, the participants were given structured diary prompts informed by the theoretical framework to record, through filling out a table, the name of the technological resource(s), the nature of the digital contents, the activities they engaged in with the resource(s), and how they interacted with the digital contents, including their attentional focus and any interaction strategies employed (See Appendix A for the diary prompts). Two example entries were provided in the diary prompts to prompt the participants to keep detailed records of their interaction with the technological resources. In order not to bias the participants towards certain behavior, examples of diversified ways of interacting with resources were provided. The participants were asked not to change their normal behaviors and to record truthfully what they did each day. They were instructed to submit a blank diary on the days they did not do anything related to English, and 72% of the participants submitted at least one blank page. The participants were instructed to keep the diary in whatever language (Chinese or English) they felt comfortable, and email their diary entries at the end of each day. Submissions were checked every night and reminders were sent the next day to those who failed to submit. Eighty-five percent of the participants submitted daily entries on time, while the rest missed one or two entries.

Interview

At the end of one month, the participants were invited to individual semi-structured interviews. The purpose of the interviews was to elicit the participants' overall perceptions of the characteristics of the technological activities that benefited vocabulary development and elaboration on their technological activities. The diary entries and the media records were used as stimuli to enhance the depth of reflection during the interviews. Interviews were conducted in Chinese through an online conferencing tool, with each interview lasting 30 to 40 minutes. During the interviews, participants were asked to give their general impressions of whether each technological activity might relate to vocabulary development, and if so, how, the nature of the digital contents, and the typical ways they interacted with these contents informed by the theoretical framework. They were also asked to rank the contributions of these technological activities to vocabulary development and elaborate on the rationale behind their ranking. We also double-checked with the participants on whether their diary entries reflected their normal English technological activities. One participant did indicate a change in her behavior as a result of participating in the study, and her data was excluded from the analysis. Memos were created immediately after each interview to summarize and reflect on the key points that had emerged in the interviews.

The vocabulary test, diary elicitation prompts, and interview protocol were pilot tested with two undergraduate students from mainland China who were not included in the main study. The instruments and the data collection procedure were revised based on the feedback.

Data Analysis

Each vocabulary test item was graded either 0 or 1 depending on whether the answer was correct. Because each derivative and word association test item contained four answers, each correct answer was given 0.25 point. The total score for the vocabulary test was 62 (M = 33.77, SD = 7.58).

The interview responses were transcribed word for word in Chinese, and the transcripts were double checked by two research assistants for accuracy. The interview responses were read through and divided into organizational segments (i.e., nature of interaction with the technological resources, and discussion of the importance of each technological resource for vocabulary development). The segments on learners' interaction with technological resources were further segmented into individual technological resources. Diary entries on the nature and ways of interaction with different technological resources were then either merged into the corresponding interview segments or added separately. Pseudonyms were used when reporting the interview responses. Content analysis was conducted on each technological resource that individual participants engaged with, using the five indicator dimensions in the theoretical framework as the coding categories (See Appendix B for the coding scheme). The three dichotomous dimensions (medium, context, and attentional focus) were coded on the two dichotomies, and values of 1 or 0 were assigned to each code to signify the presence or lack of the property (e.g., each technological experience was coded separately on the two codes of the "attentional focus" dimension–focus on meaning and focus on form. If the participant's interaction with that technological resource involved both focus on meaning and focus on form, a value of 1 was assigned to each code). The percentage of each code was calculated for each participant (e.g., percentage of focus on meaning = the total value of focus-on-meaning code divided by the total number of technological resources). Each technological resource was also coded 0, 1, or 2 on the depth of lexical information dimension to signify the lexical information that was attended to, and then an average score of this dimension across all technological resources was calculated for each participant. Following Laufer and Hulstijn (2001), each technological resource was further coded 0, 1, or 2 on each of the three sub-dimensions of the level of involvement as well as sustained engagement, and the values of all the sub-dimensions were summed for each technological resource and averaged across all technological resources to form the involvement load score for each participant. Forty-six percent of the data were re-coded independently by another researcher using the coding scheme, and the Kappa values for all the codes were above or close to 0.80, indicating strong inter-rater reliability (McHugh, 2012).

Balance scores were calculated for the three property indicators in the balance and diversification dimension of the theoretical framework (medium, context, and attentional focus) to assess the degree to which the participants' engagement in diversified technology activities was balanced. We used the percentage score of multimodal resources as the balance score of "medium" because multimodal resources denote the diversity and balance of modality (1). We devised a formula to calculate the balance score of "context" (2) and "attentional focus" (3) for each participant: subtract 1 from the absolute difference between the percentage of one code and that of the other for each indicator so that a value closer to 1 indicates greater balance (e.g., if a participant received a score of 1 for the indicator dimension "context," it indicated that the percentage of technological activities providing contextualized exposure to vocabulary engaged in by the learner was equal to that of activities providing decontextualized vocabulary study).

$$\text{Balance}_{\text{medium}} = \text{percentage}_{\text{multimodal}} \tag{1}$$

$$\text{Balance}_{\text{context}} = 1 - |\text{percentage}_{\text{contextualized}} - \text{percentage}_{\text{decontextualized}} \tag{2}$$

$$\text{Balance}_{\text{attentional focus}} = 1 - |\text{percentage}_{\text{meaning focus}} - \text{percentage}_{\text{form focus}} \tag{3}$$

Considering that learners might pay dual attention to both comprehending the text and learning the language forms when interacting with digital resources, an alternative balance score for "attentional focus" was calculated for each participant (4).

$$\text{Dual focus} = \frac{\text{\# of technological resources with dual attention}}{\text{\# of technological resources}} \tag{4}$$

The two indicators on the depth of engagement dimension–"depth of lexical information" (5) and "involvement load" (6)–indicated the depth of engagement with vocabulary. The scores for these two indicators were calculated by dividing each participant's sum score by their total number of technological resources.

$$\text{Depth of Lexical Information} = \frac{\text{Sum score of depth of lexical information}}{\text{\# of technological resources}} \tag{5}$$

$$\text{Involvement load} = \frac{\text{Sum score of involvement load}}{\text{\# of technological resources}} \tag{6}$$

In all, for each participant, six scores were calculated–four on the balance and diversification of technological experience and two on the depth of engagement–to characterize the participant's technology activities.

To answer the research questions, Pearson Correlation coefficients were calculated by correlating each participant's vocabulary score with the indicator score(s) of the participant's technology activities, both being continuous variables. The vocabulary score was close to normally distributed (Skewness = 0.23; Kurtosis = -0.96), and median split was used to divide the participants into high-achieving and low-achieving groups (DeCoster et al., 2011). As the data for quite a few independent variables were not normally distributed in the sub groups, we conducted the non-parametric alternative to an independent t-test, the Mann-Whitney U test, which does not rely on distributional assumptions, to compare the values of the property indicators of technological experience across the two groups. Accordingly, median and Interquartile range (IQR) were reported to describe the non-normally distributed data.

Thematic analysis was conducted on the interview data segments concerning the benefits of different technological resources for vocabulary development to determine potential reasons behind the relationships between different indicators and vocabulary score.

Findings

Association Between the Balance and Diversification Dimension and Vocabulary Score

The participants reported using an average of five technological resources for English learning beyond the classroom (Mdn = 5, IQR = 3; Min = 2, Max = 8). The correlation analyses revealed a negligible association between the number of digital resources and vocabulary score ($r = 0.10$, $p = 0.51$), which indicated that the number of digital resources utilized by a participant was not related to the participant's vocabulary score.

Balance of medium (i.e., multimodality) exhibited a significant positive correlation with vocabulary score ($r = 0.33$, $p < .05$), with the high achievers accessing a significantly higher proportion of multimodal resources (Mdn = 0.63, IQR = 0.25) than the low achievers (Mdn = 0.46, IQR = 0.33) ($U = 144$, $p = .008$). These findings suggested that more frequent use of technological resources using multiple modality was associated with higher vocabulary scores. In the interview responses and journal entries, some participants indicated that they felt that multimodal resources activated multiple sensory channels for learning. One participant, Chen, reported that multimodal resources (e.g., movies) "activated and combined different senses, and made [her] pick up the words faster." The participants also felt that multimodality facilitated long-term memory. They reported that "the bilingual subtitles plus the pictures left deep impressions concerning some words and expressions" (Zhang), because it enabled them to "remember the words through two levels–the picture and the printed word" (Su).

Table 2

Correlation between the Property Indicators of the Balance and Diversity Dimension and Vocabulary Score (Pearson Correlation Coefficients)

	Diversity of resources	**Indicators of the balance and diversity dimension**			**Sub-dimensions of the indicator**				
	# of digital resources	Balance of medium	Balance of context	Balance of attention focus	Dual focus	Proportion of contextualized materials	Proportion of decontextualized materials	Proportion of attention to meaning	Proportion of attention to form
Vocabulary score	0.10 (0.51)	0.33* (0.02)	-0.26 (0.09)	0.28 (0.06)	0.61** (0.00)	0.29 (0.05)	-0.22 (0.14)	0.18 (0.24)	0.44** (0.00)

Note. The values in parentheses are the *p* value of the correlation coefficient. * $p < .05$; ** $p < .01$

However, balanced access to contextualized and decontextualized materials did not show up to be a significant correlate of vocabulary score, nor did it differentiate the higher achievers from the low achievers, as shown in Table 2 and 3. Nonetheless, some participants expressed favorable perceptions of accessing contextualized materials, feeling that contextualized materials led to deeper processing and an enhanced understanding of how words should be used in different contexts. For instance, one of the participants, Ouyang, commented "When you studied a word in its context, it not only helped you remember the meaning but also helped you understand where and how it was used." Other participants attributed the benefits to the enhancement of memory. For Wang, "studying a word out of context made it feel quite distant" and "meaningless," but "accessing the word in an article or from others' narration gave it a context that was associated with either a story or a feeling, which made it easier to remember the word."

This study used two scores to examine the correlation between balanced attentional focus and technological resources. It was found that balanced exposure to technological activities with meaning focus and with form focus was not significantly associated with vocabulary score ($r = 0.28$, $p > .05$). Instead, simultaneous dual focus on both comprehending the information and learning the language when interacting with technological resources had a strong positive correlation with vocabulary score ($r = 0.61$, $p < .01$) (See

Table 2). The high achievers (Mdn = 0.50, IQR = 0.33) also had significantly greater dual attentional focus while interacting with the resources than the low achievers (Mdn = 0.33, IQR = 0.33) (U = 104, $p < .001$) (See Table 3).

The participants' interview responses and diary entries suggested that a dual focus on both the content and the language when using the digital resources was associated with higher levels of behavioral and cognitive engagement, which might have contributed to vocabulary learning. For instance, one participant, Jiang, reported that, for her, watching TV shows was both for entertainment and for learning English. Accordingly, she reported paying attention to the captions while watching, and pausing and replaying to make sure she understood each word in the captions. She recorded in her journal entries that she "watched both the English and Chinese subtitles, paid attention to the frequently occurring words, and consulted dictionaries on some words that were of interest or were critical to understanding the plot." As a result, she reported acquiring everyday words and idiomatic expressions from watching TV shows. In contrast, Chen considered watching TV shows merely as a means of relaxation and thus "seldom paid attention to the words." In her journal entries, she talked about "primarily looking at the Chinese subtitles and the pictures to follow the plot." Consequently she ranked TV shows as the least beneficial to vocabulary development among the digital resources she used.

Table 3

Nature of the Balance and Diversification Dimension across High Achievers and Low Achievers

	Dimension		***N***	**Max.**	**Min.**	**Median**	**IQR**	***U***	***p*-value**
Diversity of Resources	Number of Digital Resources	High Achiever	22	n.a.	0	5	3	221.5	.432
		Low Achiever	24			4.50	3		
Indicators of the Balance and Diversification Dimension	Balance of Multimodal Resource	High Achiever	22	1	0	0.63	0.25	144**	.008
		Low Achiever	24			0.46	0.33		
	Balance of Contextualized Material	High Achiever	22	1	0	0.31	0.50	311	.298
		Low Achiever	24			0.40	0.20		
	Balanced Attentional Focus	High Achiever	22	1	0	0.67	0.32	191	.103
		Low Achiever	24			0.47	0.26		
	Dual Attentional Focus	High Achiever	22	1	0	0.50	0.33	104**	.000
		Low Achiever	24			0.33	0.33		
Sub-dimensions of the Indicators	Proportion of Decontextualized Materials	High Achiever	22	1	0	0.15	0.27	306	.347
		Low Achiever	24			0.20	0.10		
	Proportion of Contextualized Materials	High Achiever	22	1	0	0.85	0.25	202	.165
		Low Achiever	24			0.80	0.10		
	Proportion of Attention to Meaning	High Achiever	22	1	0	0.83	0.31	210	.222
		Low Achiever	24			0.80	0.18		
	Proportion of Attention to Form	High Achiever	22	1	0	0.50	0.29	150*	.012
		Low Achiever	24			0.25	0.17		

Note. * $p < .05$; ** $p < .01$

Association Between the Depth of Engagement Dimension and Vocabulary Score

Table 4 shows that the depth of lexical information was associated positively with vocabulary score (r = 0.50, p < .01). The high achievers (Mdn = 1.00; IQR = 0.79) were found to pay significantly greater attention to words than the low achievers (Mdn = 0.67, IQR = 0.39) (U = 143, p = .013) (see Table 5).

Table 4

Correlation between the Indicators of the Depth of Engagement Dimension and Vocabulary Score (Pearson Correlation Coefficients)

	Property Indicators of the Depth of Engagement Dimension	
	Depth of Lexical Information	Involvement Load
Vocabulary Score	0.50** (0.00)	0.56** (0.00)

Note. The values in the parentheses are the p value of the correlation coefficient. ** p < .01

Table 5

Nature of the Depth of Engagement Dimension across High Achievers and Low Achievers

Indicator		***N***	**Max.**	**Min.**	**Median**	**IQR**	***U***	***p*-value**
Depth of Lexical Information	High Achiever	22	2	0	1.00	0.79	143*	.013
	Low Achiever	24			0.67	0.39		
Involvement Load	High Achiever	22	8	0	2.38	1.92	125**	.004
	Low Achiever	24			1.58	1.19		

Note. * p < .05; ** p < .01

Participants' interview responses also revealed that they tended to perceive a technological activity as not beneficial to vocabulary development if they paid no attention to word meaning and/or usage. For instance, one participant, An, ranked TV shows as the most useful resource for vocabulary learning and songs as the least helpful because "[he] felt that [he] watched TV shows to pick up colloquial expressions, listened to TED talks for exposure to different ways of thinking, thus paying less attention to language; and listened to songs for pure relaxation." Similarly, Liang, another participant, regarded public English learning channels as the most useful for vocabulary learning because he would check the meaning of the words and their usage. However, he would normally either guess the meaning of unknown words or ignore them while watching movies, and accordingly, he ranked it as the least beneficial resource for vocabulary learning. Moreover, the participants also perceived that conscious attention to word usage was more beneficial to vocabulary development than attention to word meaning only. One participant, Dai, felt that reading the news was more beneficial to vocabulary learning than vocabulary apps because in the former, he paid attention to the contexts in which words were used, whereas in the latter case, he focused only on the meaning and the pronunciation of the words and achieved "shallow understanding." The interview responses were corroborated by a strong positive association between attention to word usage and vocabulary score (r = 0.45, p < .01), but a negligible association between attending to word meaning and vocabulary score (r = 0.08) (See Table 6). Moreover, the high achievers reported paying significantly greater attention to word usage (M = 0.23, SD = 0.22) than the low achievers (M = 0.09, SD = 0.13) (U = 159, p = .016), whereas there was no significant difference in the proportion of attention to word meaning across the two groups (see Table 7). Thus, there was a strong association between attention to word usage and vocabulary score.

Table 6

Correlation between the Depth of Lexical Information and Vocabulary Score (Pearson Correlation Coefficients)

	Proportion of No Attention to Words	Proportion of Attention to Word Meaning	Proportion of Attention to Word Usage
Vocabulary Score	-0.44** (0.00)	0.08 (0.59)	0.45** (0.00)

Note. The values in the parentheses are the *p* value of the correlation coefficient. ** $p < .01$

Table 7

Nature of the Depth of Lexical Information across High Achievers and Low Achievers

		N	Max.	Min.	Median	IQR	*U*	*p*-value
Proportion of No Attention to Words	High Achiever	22	1	0	0.25	0.43	363*	.029
	Low Achiever	24			0.50	0.33		
Proportion of Attention to Word Meaning	High Achiever	22	1	0	0.38	0.30	267	.947
	Low Achiever	24			0.40	0.38		
Proportion of Attention to Word Usage	High Achiever	22	1	0	0.25	0.46	159*	.016
	Low Achiever	24			0.00	0.25		

Note. * $p < .05$

Concerning the lack of association between attention to word meaning and vocabulary score, participants felt that "checking the meaning of the words wouldn't leave deep impressions unless the words were encountered many times in other contexts" (Ouyang). The repetition of words, and the lack thereof, explained the differences in the vocabulary learning from audiovisual resources of two participants, Zhang and Duwang: Zhang reported picking up new words through checking word meaning while watching TV shows, whereas Duwang reported that checking word meaning while watching video clips did not help him accumulate vocabulary. According to Zhang, repetition was the key: "the words in sitcoms were not complex, and some expressions were repeated quite often in the shows." Paying no attention to words while interacting with digital resources had a significant negative association with vocabulary score ($r = -0.44$, $p < .01$). The high achievers also reported a significantly lower proportion of no attention to words (Mdn = 0.25, IQR = 0.43) than the low achievers (Mdn = 0.50, IQR = 0.33) ($U = 363$, $p = .029$). The identified negative association is understandable because the more the participants engaged in such experiences, the less time they spent on activities that involved conscious attention to and learning of vocabulary. The interview responses further revealed that the participants' perceptions of a technological experience shaped whether and how they attended to words. One participant, Tang, positioned movies as a way to automatize the memory of previously learned words. Consequently, she only paused the movies and checked the bilingual subtitles when she encountered words that she had learned before but couldn't recall the meaning. Similarly, another participant, Guo, positioned BBC news channel "not as a venue to learn new words but as a venue to review recently learned vocabulary." Accordingly, he "did not pay attention to the unknown words" in the news. Moreover, perceiving different affordances among technological experiences might have led to differential attention to lexical information when interacting with the resources. Jian felt that the words appearing in online news were common expressions, which made online news a valuable resource for obtaining useful information about word usage. Thus, when reading online news, she consciously focused on how words were used in particular contexts and thought about alternative ways of expression. However, she felt that a vocabulary app was meant to support systematic study of English vocabulary through repeated access. Accordingly, when interacting with a vocabulary app, she constantly reviewed the words learned in previous days before studying new words, and would not pay attention to the exemplary sentences provided in the app.

Table 4 shows that the depth to which the participants processed the words, as indicated by the involvement load score, correlated significantly with vocabulary score ($r = 0.56$, $p < .01$). Moreover, the high achievers (Mdn = 2.38, IQR = 1.92) had a significantly higher level of involvement load with vocabulary than the low achievers (Mdn = 1.58, IQR = 1.19) ($U = 125$, $p = .004$) (See Table 5). The findings suggest that the deeper the participants engaged with the words, the greater their vocabulary score might be. The depth of processing was influenced by perceived agency during the engagement. One participant, Zhou, compared her experience of reading the news and studying through a vocabulary app:

> When interacting with the vocabulary app, I played a more passive role; whereas when reading news, my roles were more active. Because I was more active, I would pay attention to the language and actively link and contrast the word with related words and associate it with sentences, and thus get a deeper impression, whereas, for the vocabulary app, it was mainly about rote memorization.

The participants' perceptions of the value of the language they obtained from their experiences also influenced their involvement load. An said that she did not process the vocabulary in songs because the language was not useful: "I did not consciously understand nor remember the words because many expressions in the lyrics were not the standard use". However, she put more effort into understanding the words that appeared in essays because they "were actually being used" and were thus "more useful."

Discussion

Understanding how learners' out-of-class technology-mediated language activities are associated with L2 vocabulary knowledge is an important issue. Previous studies have primarily approached this issue from the lens of individual technological resources and suggested possible characteristics that underlie their language learning potential (Arndt & Woore, 2018; De Wilde et al., 2020; Peters, 2018). Instead of focusing on individual resources, this study examined a group of EFL learners' overall out-of-class technological experience, developed a framework of indicators to gauge the quality of their digital activities, and examined how these indicators might be associated with their L2 English vocabulary score. The findings revealed that indicators along both the balance and diversification dimension and the depth of engagement dimension related significantly to the participants' vocabulary score. The findings supported theoretical propositions concerning the critical conditions for vocabulary learning (Laufer, 2017; Schmitt, 2014). The findings suggest that both the diversity of the experience and the depth of engagement with vocabulary during the experience deserve attention in discussions on the relationship between out-of-class technological experience and vocabulary knowledge. Given the exploratory nature of this correlational study, the findings are suggestive at most, and future research may examine the nature of the causal relationship between these characteristics and vocabulary knowledge.

The study found that access to technological resources that activated both audiovisual and textual modality showed a significant positive association with vocabulary scores. This result is consistent with findings from previous experimental studies (Perez et al., 2013), and suggests that multimodality might be an important consideration in naturalistic out-of-class technological use as it facilitates the establishment of form-meaning links (De Wilde et al., 2020; Peters, 2018). Contrary to the hypothesis, balanced access to contextualized and decontextualized learning materials beyond the classroom did not correlate significantly with learners' vocabulary score, and the participants valued the vocabulary learning potential of contextualized learning experiences. This finding might be biased by the particular EFL learning context, in which form-focused instruction dominated. In such contexts, more contextualized experience outside the classroom might help achieve a greater balance between incidental and deliberate learning opportunities in the overall language learning experience (Godwin-Jones, 2018; Toffoli & Sockett, 2010). The findings might be different in other learning contexts.

This study also found that the balance between meaning-focused and form-focused learning experiences was a significant correlate of vocabulary score. However, this balance did not lie in the combination of meaning-focused and form-focused learning activities, as Lee (2019) suggested, but rather in the simultaneous attention to both meaning and form during the activities. The inconsistency with Lee's study might be due to the different vocabulary knowledge measurements used; this study utilized a composite measure of both vocabulary size and vocabulary depth, whereas Lee (2019) measured vocabulary size only. More research is needed to determine how the nature of out-of-class technological experience might be related to different vocabulary measures. The finding that the degree of dual attention to both meaning and form differentiate high achievers and low achievers is consistent with previous research findings showing that combined meaning and form attention to vocabulary leads to greater learning and retention than attention to meaning or attention to form only (Karami & Bowles, 2019; Schmitt, 2008). Similarly, the study found that the depth of the lexical information the participants attended to and their levels of involvement with vocabulary encountered in their technological experiences were associated positively with the participants' vocabulary score. The findings support the proposition that greater attention and depth of processing are associated with greater vocabulary learning (Leow, 2018; Schmidt, 2010). The greater attention and depth of processing observed among the high achievers might suggest that the participants with greater vocabulary score were more likely to free up attentional resources to attend to both meaning and form and process the vocabulary deeper. It might also suggest that dual attention to both meaning and form and deeper processing of the vocabulary might be a critical condition for effective vocabulary learning. Whether dual attention and deeper processing is a necessary condition for vocabulary development or an intermediary outcome of vocabulary knowledge in the context of informal language learning with technological resources needs further exploration.

Conclusion

This study examined the relationship of a group of Chinese undergraduate students' out-of-class English language learning experiences with technology and their L2 English vocabulary score. The study found that how diversified the experiences were and how deeply learners engaged with the vocabulary during the experiences related positively to their vocabulary scores. Specifically, access to multimedia resources, dual attention to both meaning and form, and deeper processing of the vocabulary correlated significantly with vocabulary scores. The findings, although tentative due to the exploratory nature of the study, suggest that these indicators may be directions to pursue when researching the quality of out-of-class language learning with regard to vocabulary development. Educators may also raise learners' attention to these dimensions to help them make informed decisions on out-of-class digital choices and behaviors.

This study was limited in its scale and the generalizability of its findings. First, the findings might be biased by the particular research context: the dominant form-focused instruction in the teaching context, the participants being of relatively high proficiency levels, and the language being studied, English, being indicative of high social power. Language proficiency plays an essential role in incidental learning from naturalistic resources (Peters, 2018), and the social power of the language being studied may shape learners' engagement in out-of-class learning (De Wilde et al., 2020). Thus, future research may expand the research inquiry to different learner populations and language learning contexts. Second, the findings were based on correlational analysis of coded open-ended naturalistic data and vocabulary score, and comparative analyses of groups artificially formed through medium split of vocabulary score. The validity of findings depended critically on the thoroughness of the participants' self-report and the reliability of the coding. Although measures were taken to enhance the thoroughness, truthfulness, and depth of the self-report data, and interrater reliability was checked, the inherent limitations of self-report data and subjective coding make the findings exploratory at most. Although median split is a commonly used method, the artificially formed high achieving and low achieving groups through median split may reduce the ability to detect relations and lose statistical power (DeCoster et al., 2011). Despite the limitations, this study paves the way for future large-scale survey studies by suggesting some indicators to be further statistically tested with regard to their relative importance and contributions to vocabulary development across different proficiency levels. Moreover, as the participants reported different amounts of technological resources, the percentage of the indicators might have been inflated for participants who reported fewer technological resources. However, since the high achieving and low achieving groups showed no significant difference in the number of technological resources they used, the potential inflation may not have biased the findings.

Moreover, as the study intended to examine the participants' general vocabulary proficiency, a combination of measures assessing a small number of target items rather than a validated test was used. The lack of robustness of the test made the findings rather tentative. Future research may adopt a validated test assessing a particular type of vocabulary knowledge in order to obtain fine-tuned insights into the link between technological experiences and vocabulary knowledge. Moreover, the vocabulary test was administered before the diaries and interviews to help recruit participants with varying vocabulary proficiency. This arrangement may have compromised the association between the participants' diary entries and their vocabulary performance. However, the diary study was conducted immediately after the vocabulary test, and it was unlikely that students' vocabulary performance would change drastically within one month. Moreover, measures were taken to increase the likelihood that diary entries would reflect the participants' normal learning behaviors. Nonetheless, the research findings need to be interpreted with caution. Future studies may include another round of vocabulary tests immediately after the diary and interview study. Moreover, a longitudinal approach might help unravel the dynamic interaction between different dimensions of out-of-class technological experience and vocabulary knowledge at different stages of development.

Acknowledgements

This study was funded by the Hong Kong General Research Fund (No. 17605518).

References

Al-Surmi, M. (2012). Authenticity and TV shows: A multidimensional analysis perspective. *TESOL Quarterly*, *46*(4), 671–694. http://dx.doi.org/10.1002/tesq.33

Ala-Kyyny, J. (2012). *The role of English-language music in informal learning of English* [Unpublished master thesis]. University of Jyväskylä. https://jyx.jyu.fi/bitstream/handle/123456789/38449/URN:NBN:fi:jyu-201209032302.pdf;sequence=1

Arndt, H. L., & Woore, R. (2018). Vocabulary learning from watching YouTube videos and reading blog posts. *Language Learning & Technology*, *22*(3), 124–142. http://128.171.57.22/bitstream/10125/44660/1/22_03_arndt_10125-44660.pdf

Benson, P. (2011). Language learning and teaching beyond the classroom: An introduction to the field. In P. Benson & H. Reinders (Eds.), *Beyond the language classroom* (pp. 7–16). Palgrave Macmillan.

Blyth, D. A., & LaCroix-Dalluhn, L. (2011). Expanded learning time and opportunities: Key principles, driving perspectives, and major challenges. *New Directions for Youth Development*, *131*, 15–27. http://dx.doi.org/10.1002/yd.405

Clothey, R. A. (2012). Globalization, decentralization and equity: Post-secondary English language policy in China. In C. Acedo, D. Adams, & S. Popa (Eds.*), Quality and qualities: Tensions in education reforms (pp. 163–182)*. Sense Publishers.

Cole, J., & Vanderplank, R. (2016). Comparing autonomous and class-based learners in Brazil: Evidence for the present-day advantages of informal, out-of-class learning. *System*, *61*, 31–42. http://dx.doi.org/10.1016/j.system.2016.07.007

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2020). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition*, *23*(1), 171–185. http://dx.doi.org/10.1017/S1366728918001062

DeCoster, J., Gallucci, M., & Iselin, A. M. R. (2011). Best practices for using median splits, artificial categorization, and their continuous alternatives. *Journal of experimental psychopathology*, *2*(2), 197–209. http://dx.doi.org/10.5127/jep.008310

Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition*, *42*(3), 499–523. http://dx.doi.org/10.1017/S0272263119000494

Garnier, M., & Schmitt, N. (2016). Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System*, *59*, 29–44. http://dx.doi.org/10.1016/j.system.2016.04.004

Godwin-Jones, R. (2018). Contextualized vocabulary learning. *Language Learning & Technology*, *22*(3), 1–19. https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/fe8e9343-3e0c-4bfb-858f-c47c4f36ff41/content

Hou, J. (2017). *The dynamics of English writing development in advanced Chinese learners.* University of Groningen.

Hui, D. (2004). Reflection on vocabulary size of Chinese university students. *International Education Journal*, *5*(4), 571–581.

Hulstijn, J. H. (2013). Incidental learning in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics (Vol. 5, pp. 2632–2640).* Wiley-Blackwell.

Karami, A., & Bowles, F. A. (2019). Which strategy promotes retention? Intentional vocabulary learning, incidental vocabulary learning, or a mixture of both? *Australian Journal of Teacher Education*, *44*(9), 25–43.

Lai, C. (2015). Perceiving and traversing in-class and out-of-class learning: Accounts from foreign language learners in Hong Kong. *Innovation in Language Learning and Teaching*, *9*(3), 265–284. http://dx.doi.org/10.1080/17501229.2014.918982

Lai, C. (2018). *Autonomous language learning with technology beyond the classroom.* Bloomsbury Publishing.

Lai, C. (2019). Learning beliefs and autonomous language learning with technology beyond the classroom. *Language Awareness*, *28*(4), 291–309. http://dx.doi.org/10.1080/09658416.2019.1675679

Lai, C., Hu. X., & Lyu, B. N. (2018). Understanding the nature of learners' out-of-class language learning experience with technology. *Computer Assisted Language Learning*, *31*(1), 114–143. http://dx.doi.org/10.1080/09588221.2017.1391293

Laufer, B. (2017). From word parts to full texts: Searching for effective methods of vocabulary learning. *Language Teaching Research*, *21*, 5–11. http://dx.doi.org/10.1177/1362168816683118

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1–26. http://dx.doi.org/10.1093/applin/22.1.1

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 33–51. http://dx.doi.org/10.1177/026553229901600103

Lee, J. S. (2019). Informal digital learning of English and second language vocabulary outcomes: Can quantity conquer quality? *British Journal of Educational Technology, 50*(2), 767–778. http://dx.doi.org/10.1111/bjet.12599

Leow, R. P. (2018). Noticing hypothesis. In J. I. Liontas & M. DelliCarpini (Eds.), *The TESOL encyclopedia of English language teaching* (pp. 1–7). Wiley-Blackwell.

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. http://dx.doi.org/10.1080/14790718.2012.679275

Lyrigkou, C. (2019). Not to be overlooked: Agency in informal language contact. *Innovation in Language Learning and Teaching, 13*(3), 237–252. http://dx.doi.org/10.1080/17501229.2018.1433182

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia medica, 22*(3), 276–282. http://dx.doi.org/10.11613/BM.2012.031

Nation, I. S. P. (1990). *Teaching and learning vocabulary.* Heinle and Heinle

Nation, I. S. P. (2012). *The vocabulary size test: Information and specifications*. Victoria University of Wellington. https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-tests/the-vocabulary-size-test/Vocabulary-Size-Test-information-and-specifications.pdf

Nation, I.S. P. (2015). Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language, 27*(1), 136–145.

Perez, M. M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System, 41*(3), 720–739.

Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL-International Journal of Applied Linguistics, 169*(1), 142–168. http://dx.doi.org/10.1075/bct.109.itl.00010.pet

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in language assessment* (pp. 41–60). Lawrence Erlbaum.

Richards, J. C. (2015). The changing face of language learning: Learning beyond the classroom. *RELC Journal, 46*, 5–22. http://dx.doi.org/10.1177/0033688214561621

Rodgers, M. P., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly*, *45*(4), 689–717.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010, Singapore, December 2–4* (pp. 721–737). National University of Singapore, Centre for Language Studies.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning, 64*(4), 913–951. http://dx.doi.org/10.1111/lang.12077

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching, 52*(2), 261–274. http://dx.doi.org/10.1017/S0261444819000053

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, *36*(2), 145–171. http://dx.doi.org/10.2307/3588328

Steel, C. H., & Levy, M. (2013). Language students and their technologies: Charting the evolution 2006–2011. *ReCALL*, *25*(3), 306–320. http://dx.doi.org/10.1017/S0958344013000128

Sundqvist, P. (2011). A possible path to progress: Out-of-school English language learners in Sweden. In P. Benson & H. Reinders (Eds.), *Beyond the language classroom* (pp.106–118). Palgrave Macmillan.

Sundqvist, P. (2019). Commercial-off-the-shelf games in the digital wild and L2 learner vocabulary. *Language Learning & Technology*, *23*(1), 87–113. https://scholarspace.manoa.hawaii.edu/bitstream/10125/44674/1/23_01_10125-44674.pdf

Sundqvist, P., & Sylvén, L. K. (2016). *Extramural English in teaching and learning: From theory and research to practice*. Palgrave Macmillan.

Toffoli, D., & Sockett, G. (2010). How non-specialist students of English practice informal learning using web 2.0 tools. *ASp. la revue du GERAS, 58*, 125–144. http://dx.doi.org/10.4000/asp.1851

Vanderplank, R. (2019). 'Gist watching can only take you so far': Attitudes, strategies and changes in behavior in watching films with captions. *The Language Learning Journal, 47*(4), 1–17. http://dx.doi.org/10.1080/09571736.2019.1610033

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*(1), 219–258. http://dx.doi.org/10.1111/j.1467-9922.2010.00593.x

Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 159–168). Routledge.

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). Routledge.

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Appendix A. The Diary Prompts

Please fill in the technological resources or activities you've engaged in that are in English or give you access to English today. Please see the following examples to see what to write about for each technological resource or activity.

Example:

Technological resources / tools / activities	**When and where did I use it?**	**What did I use?**	**How long did I use it?**	**How did I use it?**
I listened to English songs	On the way to school (MTR)	I listened to English songs on my phone. The songs I listened to were from Allen Walker, the Lincoln Park, and Adele.…	35 minutes	My attention was on the rhythms. I didn't look at the lyrics while listening. I sang along (silently) as I was quite familiar with these songs and had listened to them several times
I read and watched BBC news report	On the way to school and back home In the evening, on the desk of my room	I went to the BBC website to read some latest news. I read a news report on the earthquake in Indonesia, and a news report on China-US trade war. I also read a news report on how a rover on the Mars that stopped working and watched the video.	1 hour (20 minutes on the MTR; 40 minutes at home)	When I read the news on BBC on MTR, I focused on getting the gist. But I searched for the meaning of a few words on the Google Translator, and read aloud a few phrases. At home, I read the news several times since they were interesting but hard to understand. I checked the meaning of a few words in my dictionary app. I read the subtitles while watching the videos.

(Note: The examples show the details need to record in your diary. They are not behaviors for you to imitate. Please truthfully record the resources you used and how you interacted with the resources. Please be as detailed as possible)

Your Diary:

Technological resources / tools / activities	**When and where did I use it?**	**What did I use?** **(topics; content)**	**How long did I use it?**	**How did I use it?** **(Attentional focus? Did anything with the words in the material?)**

Appendix B. The Coding Scheme

Indicator dimension	Code	Interrater reliability (kappa)	Value	Definition & Example	Notes
Medium (Whether the processing of the content involved both audio-visual and textual modality)	Multimodal	0.81	1: Yes 0: No	Used both audio or visual and textual information when interacting with the material (e.g., "I checked the captions while watching the TV shows")	Reading apps and video resources without captions or subtitles were coded as single; Both checking the captions (L2) and subtitles (L1) while watching or listening to resources were coded as multimodal;
	Single	0.85	1: Yes 0: No	Primarily relied on only one modality (either audio/visual or textual) when interacting with the material (e.g., "I wouldn't read the lyrics while listening to songs")	Online news may be coded as either single or multimodal depending on whether the participants accessed the accompanying video
Context (Whether the vocabulary exposure was contextualized or used in a contextualized way)	Contextualized	0.93	1: Yes 0: No	Vocabulary items were by nature embedded in contexts (e.g., TV shows, reading apps, listening apps, news, academic paper, etc. were coded as contextualized by default)	Public English language channels may contain both contextualized and decontextualized materials, and thus may both be coded as based on the participants' description of the content
	Decontextualized	0.91	1: Yes 0: No	Vocabulary items were removed from their context to be focused on as language items (e.g., vocabulary app were coded as decontextualized by default)	In cases where the participants did not say whether the materials were contextualized or not, the materials were coded by its nature (e.g., news, movies, etc. were coded as contextualized by default)
Attentional Focus (Whether the participants paid primary attention to meaning or to language form)	Attention to meaning	0.86	1: Yes 0: No	Primary attention on comprehending/enjoying the content when interacting with the resource (e.g., "When I watched the movies, my attention was on the story. I primarily followed the Chinese subtitles")	If an individual participant indicated having both intentions when interacting with a resource, then that resource was coded both (e.g., watching TV shows for the storyline but consciously paying attention to idioms and expressions used in the shows by checking the captions; pausing the video or replaying the video to check on some expressions)
	Attention to form	0.77	1: Yes 0: No	Primary attention on studying language form when interacting with the resource (e.g., "I read the Chinese subtitles while listening to the conversation in the movie. I try to match the English words I heard with the Chinese subtitle")	
Depth of Lexical Information	No attention to words	0.81	0	No attention to words in the resource (e.g., " I seldom listened to what the words are when listening to	If an individual reported watching videos without looking at the caption/subtitles, or paying

Indicator dimension	Code	Interrater reliability (kappa)	Value	Definition & Example	Notes
(What lexical information did the participant pay attention to)				songs")	attention to the rhythms of the songs, or reading academic papers by translating blocks of texts and skimming through the translated text only, it was coded as no attention to words.
	Conscious attention on word form-meaning link only		1	Paying attention to the meaning of the word only (e.g., "The TV shows had both English and Chinese subtitles. When I watched the TV shows, I'd look at the subtitles and consciously compare the English words against the Chinese subtitles")	If an individual checked the glosses of words, dictionaries, or checked the Chinese subtitles of individual words while reading or watching, it was coded as attention to form-meaning link.
	Conscious attention on word usage		2	Paying attention to the multiple meanings of a word, the synonyms, or the usage of a word (e.g., "I usually paid attention to the phrases, some interesting idiomatic expressions, and the usage of words")	If an individual mentioned associating words with related words, noticing the different meaning and usage of familiar words, paying attention to which words went with which words, it was coded as attention to word usage
Involvement Load (adapted from Laufer & Hulstin, 2001) (How deeply did the participant process the word?)	Need	0.81	0	No motivation to look for or use the words (e.g., "I seldom looked at the English captions while watching TV")	For reading apps that contain a glossary at the end, if the participant studied the words in the glossary after finishing the text, it was coded as initiated by technological resource; and if the participant self-initiated to check unknown words encountered either against the glossary or the dictionary, it was coded as initiated by the learners. The vocabulary apps and the public English language channels that focused on word explanation would be automatically coded as word study initiated by technological resource
			1	Motivation to study the words initiated by the technological resource (e.g., "The vocabulary app provided an exemplar sentence and asked us to guess the meaning of the word from the context first and then showed us the meaning of the word")	
			2	Motivation to study the word initiated by the learners (e.g., "I focused on the English caption while watching")	
	Search	0.76	0	Didn't search the word (e.g., "Sometimes I'd hear some strange usage. But I'd just let it go")	If the participant searched the dictionary or the Chinese subtitles/lyrics on how a word is used in the context, then it was coded searched the form/usage
			1	Searched for the meaning by checking the dictionary, the glossary, or the Chinese subtitles/lyrics (e.g., "I would check the meaning of the words that I was interested in")	

Indicator dimension	**Code**	**Interrater reliability (kappa)**	**Value**	**Definition & Example**	**Notes**
			2	Searched for the form or usage (e.g., "Sometimes when I was not sure how to express, I would go check how to use that particular word")	
	Evaluation	0.78	0	No evaluation was involved (e.g., "When I met new words, I turned to the Chinese subtitles for their meanings and then continued watching")	1 always involved evaluating the specific meaning of the word in the particular context against the possible meanings of the word in language input activities;
			1	The meaning or usage of words are compared with other words in provided contexts (e.g., "I checked the meanings of the word in the dictionary and then switched back to the app to see what specifically the meaning of the word was in that particular sentence")	2 always involved language output activities (self-generated English translation; email; etc.)
			2	The meaning or usage of words are compared with other words in self-generated sentences (e.g., "The movie only had Chinese subtitles. I would see the Chinese text and then think what the English would be. When I found the English I came up with were exactly the same as what I heard, I would have a sense of accomplishment")	
	Sustained engagement beyond the moment	0.79	0	Did not take any notes for later processing (e.g., "I wouldn't take down the words on the notebook. That would break the flow of the movie")	If the participant took screen captures of words and save on the computer but didn't go back to review them later, it was also coded 1
			1	Took notes on or saved the words encountered (e.g., "Sometimes I met some new words. I'd search online and then add them to the notebook")	
			2	Reviewed words studied before or recorded in the notebook or consciously used the words in subsequent communicative	

Indicator dimension	Code	Interrater reliability (kappa)	Value	Definition & Example	Notes
				contexts (e.g., “I copied them down on the notebook and reviewed all the words at the end of the week”)	