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

2026, Volume 30, Issue 1
pp. 1–26

ARTICLE

Enhancing college students' listening performance and self-efficacy using mobile annotated vlogs

Hui-Tzu Hsu, Center for Language Studies, National Chung Cheng University

Abstract

Previous research has demonstrated that annotated or captioned videos can enhance second language learners' listening comprehension and vocabulary acquisition. However, the influence of annotated vlogs delivered through mobile platforms on these outcomes remains underexplored. Accordingly, this study investigated the effects of annotated vlogs delivered through mobile devices on English as a foreign language learners' listening comprehension, vocabulary retention, and self-efficacy. This quasi-experimental study recruited 131 college students at a pre-intermediate level of English proficiency, enrolling 43 in the control group that learned through audio files, 44 in the first experimental group that learned through captioned vlogs, and 44 in the second experimental group that learned through annotated vlogs. The research instruments comprised listening pre- and post-tests, vocabulary pre- and post-tests, initial and post-intervention surveys evaluating self-efficacy, a learning perception survey for the annotated vlog group, and semi-structured interviews with members of the two experimental groups. The results revealed that students learning with annotated vlogs significantly outperformed those in the other two groups in listening comprehension and vocabulary retention. The learners in the annotated vlog group also reported the greatest post-intervention increase in self-efficacy.

Keywords: mobile listening learning; annotated vlog; language performance; self-efficacy; learning perceptions

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

APA Citation: Hsu, H. T. (2026). Enhancing college students' listening performance and self-efficacy using mobile annotated vlogs. *Language Learning & Technology*, *30*(1), 1–26. https://doi.org/10.64152/10125/73691

Introduction

Although listening is a vital L2 skill, it is difficult to practice (Vandergrift, 2004) and often marginalized in instruction (Newton & Nation, 2020). Learners struggle with rapid speech, unfamiliar vocabulary, and accents (Vandergrift & Goh, 2012), and traditional practices such as repeating audio, reading scripts while listening, or pausing to identify new words often fail to address these challenges (Chao, 2013). Annotated videos with L1 keyword glosses or highlighted English captions can improve listening comprehension and vocabulary learning (Montero Perez et al., 2018). However, these tools typically position learners as passive recipients and offer limited interaction or feedback, and their delivery through mobile technologies remains underexplored (Çakmak & Erçetin, 2018).

A vlog (video blog) refers to a form of blogging in which messages are created and communicated through video (Merriam-Webster, 2022). In language learning, vlogs support communication and collaborative meaning-making (Wood, 2019). Unlike conventional platforms such as YouTube, which often provide unsystematic instruction, limited interaction, and subtitles with little pedagogical support (Sila & Gunawan, 2025), vlogs can offer more structured learning environments. Although vlog research has mainly focused on speaking and presentation skills (Zhang et al., 2021), some studies also indicate its potential to improve listening comprehension and vocabulary learning (Aldukhayel, 2021a, 2021b; Arndt

& Woore, 2018).

Mobile technology further enhances this potential by enabling authentic and flexible learning beyond the classroom (Hao et al., 2019). Accordingly, this study examined annotated vlogs accessed through mobile phones to address listening difficulties and support vocabulary retention among L2 learners (Chen et al., 2022). The instructor-produced vlogs were delivered through Padlet, a mobile-friendly collaborative platform for video sharing and commenting. Unlike standard vlogs or captioned videos, they included pedagogical annotations highlighting target vocabulary with images and Chinese-English definitions and allowed learner-controlled replay in an interactive environment. This multimodal support facilitated form-meaning connections and reduced cognitive load, thereby supporting listening comprehension, vocabulary retention, and awareness of learning progress (Mayer & Moreno, 2003).

Although vlogs can support listening comprehension (Aldukhayel, 2021a) and vocabulary learning (Arndt & Woore, 2018), several gaps remain. First, research rarely examines annotated vlogs delivered through mobile technologies. Existing studies instead focus on video captions and mobile learning motivation (Hsu, 2015), multimedia annotation for text recall and incidental vocabulary learning (Çakmak & Erçetin, 2018), or annotated videos integrated with vocabulary systems (Chen et al., 2022). Second, studies on vlogs for listening and vocabulary learning mainly compare caption types or use noncaptioned vlogs (Aldukhayel, 2021b; Arndt & Woore, 2018), leaving annotated vlogs underexplored. Third, although self-efficacy has been studied in podcasting and instructional or short-form videos (Ragazou & Karasavvidis, 2023; Ye et al., 2025), research on teacher-produced vlogs focuses largely on attitudes and outcomes. Notably, Kay (2012) reported no empirical studies on self-efficacy in vlog-based learning. To address these gaps, this study examined Padlet-based annotated vlogs delivered via mobile phones and their effects on listening comprehension, vocabulary retention, self-efficacy, and learning perceptions.

Literature Review

Effects of Annotated Videos on Listening Comprehension and Vocabulary Acquisition

Annotation refers to glosses that support L2 reading (Nation, 2001) and may also facilitate incidental vocabulary learning in listening (Hulstijn, 2001). In computer-assisted language learning, multimodal video inputs such as sound, images, and captions support listening comprehension and vocabulary learning (Chen & Chen, 2021; Teng, 2022). Captions and annotated keywords using typographical cues, L1 glosses, or images can further aid lexical acquisition (Hsu, 2018; Sydorenko, 2010). From a cognitive load perspective, annotations guide attention and help learners process key information efficiently (Mayer & Moreno, 2003).

Although annotations consistently improve vocabulary retention (Montero Perez et al., 2018), their effects on listening comprehension are mixed. Some studies report benefits of annotated captions (Hsu et al., 2013; Winke et al., 2010), whereas others suggest multimodal input may cause cognitive overload (Aldera & Mohsen, 2013; Montero Perez et al., 2018). Nevertheless, annotations direct attention to unfamiliar words and strengthen form-meaning connections. For instance, Hsieh (2020) found L1 annotated captions improved vocabulary but not listening comprehension, while Jones and Plass (2002) showed textual-pictorial annotations produced the greatest immediate vocabulary gains and listening recall, with pictorial annotations supporting stronger long-term retention.

Mobile Listening and Vocabulary Learning

The COVID-19 pandemic accelerated the use of mobile technologies in language learning. Devices such as smartphones, tablets, and laptops are widely used in EFL instruction due to their flexibility, ubiquity, and authenticity (Loewen et al., 2019). By extending formal instruction through informal media, mobile technology enables learning anytime and anywhere (Sung et al., 2016). Research has confirmed its effectiveness across language skills, including vocabulary (Okumuş Dağdeler, 2023), speaking and listening (Li, 2024), reading (Hou et al., 2024), and writing (Zaniar et al., 2025).

However, research on multimedia annotated videos in mobile learning is limited and findings are mixed. Çakmak and Erçetin (2018) found that video annotations improved vocabulary retention but not listening comprehension. Similarly, Lin and Yu (2017) reported that combined aural, textual, and pictorial input improved word recall, reduced cognitive load, and was positively perceived. These findings suggest that mobile-assisted listening can support vocabulary learning through short, repeated, and learner-controlled activities. Nevertheless, the use of annotated vlogs for vocabulary development and listening comprehension in mobile learning remains underexplored. The present study addresses this gap by examining vlog-based activities delivered exclusively through mobile phones.

Vlog-Assisted Language Learning

Vlogs allow users to share experiences and content, with interaction mainly occurring on platforms such as YouTube and TikTok, where viewers watch, comment, and respond (He et al., 2025). Characterized by authenticity, efficiency, and accessibility (Wood, 2019), they cover diverse topics and offer flexible opportunities for language engagement beyond the classroom (Zhang et al., 2021). From Lev Vygotsky's social constructivist perspective (Vygotsky, 1962), vlogs delivered via Padlet can foster interaction, peer collaboration, and shared meaning-making (Huang, 2021). By supporting small online learning communities, vlogs encourage feedback, perspective sharing, and goal setting, thereby enhancing engagement and supporting listening practice, vocabulary development, and broader language learning (Hsieh, 2020).

Vlogs may provide useful L2 listening input and support vocabulary learning, but research on their effectiveness remains limited (Aldukhayel, 2021a). Arndt and Woore (2018) found that blogs and vlogs facilitated incidental vocabulary learning: blogs supported orthographic knowledge, whereas vlogs promoted grammatical and semantic knowledge. Consistent with dual coding theory (Paivio, 1990), blogs reinforce written forms through repeated exposure, while vlogs combine spoken language and visuals to support semantic and grammatical processing. Aldukhayel (2021b) examined vlog captions with 96 learners in L2 caption, L1 caption, and no-caption conditions. Results showed that L2 captions benefited higher-proficiency learners, while lower-proficiency learners struggled with simultaneous visual, auditory, and textual input; strong visual support improved comprehension, and L2 captions were perceived as more helpful. These findings highlight the pedagogical potential of vlog-based tasks and motivate the researcher to use instructor-produced annotated vlogs to support vocabulary learning and listening comprehension beyond caption-based approaches (Aldukhayel, 2021b). From a cognitive load theory perspective (Chandler & Sweller, 1991), captioned or annotated vlogs may reduce processing demands and facilitate form-meaning mapping, though research on annotated vlogs in mobile learning remains scarce.

Self-Efficacy

Self-efficacy, defined as individuals' belief in their ability to accomplish tasks, is central to social cognitive theory (Bandura et al., 1997). It strongly predicts persistence, engagement with challenging tasks, and academic performance, whereas low self-efficacy often leads to avoidance (Bandura & Schunk, 1981). In L2 learning, it is closely associated with learning outcomes and self-regulatory strategies (Bai et al., 2019; Chen & Lin, 2009).

Research has also examined self-efficacy across language skills. Wang et al. (2017) conceptualized EFL learners' self-efficacy in speaking, listening, reading, and writing, showing that listening self-efficacy was positively related to listening proficiency and learners' ability to process spoken input (Mills et al., 2006). Studies further indicate that self-efficacy supports listening comprehension and engagement with spoken input (Tabrizi & Saeidi, 2015) and contributes to vocabulary learning through sustained engagement (Oyama, 2022). While instructional design and multimedia support can strengthen self-efficacy by managing cognitive load and enhancing learner control (Zheng et al., 2009), research on technology-mediated listening shows that multimodal input may improve comprehension without necessarily increasing self-efficacy (Zareian et al., 2015). Consequently, the role of self-efficacy in multimedia

listening environments, particularly in relation to annotated vlogs and self-paced listening, remains underexplored.

Present Study

Previous studies have largely overlooked annotated vlogs in listening and vocabulary learning. Based on cognitive load theory (Mayer & Moreno, 2003), this study developed annotated vlog activities using signaling, alignment, and synchronization strategies. Target words were highlighted with bilingual and visual annotations to direct attention, reduce information switching, and support deeper learning. The study examined the effects of annotated vlogs on EFL learners' listening comprehension, vocabulary retention, self-efficacy, and learning perceptions. Annotated vlog tasks were delivered through Padlet.com, where students used mobile devices to view vlogs, answer comprehension questions, and provide feedback. This study addressed the following research questions:

1. Do annotated vlogs significantly influence listening comprehension?

2. Do annotated vlogs significantly influence vocabulary retention?

3. Do annotated vlogs significantly influence self-efficacy?

4. What are learners' perceptions of the benefits of annotated vlogs to listening comprehension and vocabulary retention?

It should be noted that the survey data were collected only from the annotated vlog group because the fourth research question focused on the annotation-enhanced intervention.

Methods

Participants

The study involved 131 first-year non-English-major students (aged 18-19) from three English classes at a national university in Taiwan, all at the A2 level of the Common European Framework of Reference (CEFR). The classes were randomly assigned to a control group (N = 43; men = 22, women = 21), a captioned vlog group (N = 44; men = 39, women = 5), and an annotated vlog group (N = 44; men = 15, women = 29). All groups were taught by the same instructor using standardized procedures to ensure instructional consistency.

Research Design

A quasi-experimental design was used. All groups studied the same intermediate listening materials, differing only in delivery format (see Appendix A). The captioned and annotated vlog groups accessed vlogs via Padlet on mobile devices, allowing response submission, peer feedback, instructor monitoring, and learner-controlled playback.

Vocabulary knowledge was defined as aural lexical knowledge (ALK). Matthews et al. (2024) defined ALK as "the capacity to map the phonological composition of words (form) that are encoded in speech onto an appropriate semantic representation (meaning)" (p. 65). Captioned vlogs provided English captions, whereas annotated vlogs highlighted target words with images and Chinese–English definitions. While captions offer dual-channel input, annotations focus attention on key vocabulary and may reduce cognitive load (Mayer & Moreno, 2003). The study hypothesized that annotated vlogs would enhance incidental vocabulary learning, listening comprehension, and self-efficacy. The control group used instructor-delivered audio through conventional listening activities.

The instructional materials followed the format of the TOEIC Listening Test and consisted of short conversations and talks lasting approximately 1–1.5 minutes. These discourse-level tasks were selected because they require more complex listening comprehension than picture-description or question-response items. The study included three instructional conditions: captioned vlogs, annotated vlogs, and

an audio-only control condition. Captioned vlogs provided English captions, whereas annotated vlogs included synchronized textual and visual annotations of target vocabulary. Both vlog conditions allowed learners to control playback, access multimodal support, and interact with peers through Padlet. Annotated vlogs differ from captioned vlogs in several features: (a) captioned vlogs included English captions without learner control over on/off switching of captions, and (b) annotated vlogs included English and Chinese textual annotations, visual annotations (images), no on/off control of annotated captions, and annotations that were synchronized with speech, placed above spoken words, and briefly displayed.

The four-week intervention required the two vlog groups to watch two short-conversation and two short-talk vlogs weekly on mobile devices via Padlet, each followed by three comprehension questions. The control group listened to the same materials in audio form without video, captions, and annotations. All groups completed one 50-minute session per week, including 15 minutes of listening and 35 minutes of comprehension questions and discussion. The vlog groups engaged in peer discussion and feedback on Padlet, whereas the control group participated in instructor-led discussion. Semi-structured interviews were conducted after the intervention to examine students' learning experiences.

Vlog Task Design

Over the course of 4 weeks, two treatment levels (captioned and annotated) and one control condition viewed a total of eight vlogs featuring short conversations and eight vlogs featuring short talks. The experimental groups completed mobile vlog-based tasks in captioned or annotated formats (Figures 1–4). The vlogs were created with DaVinci Resolve and uploaded to Padlet using materials from the TOEIC practice book, *TOEIC Upgrade* (Anderson, 2017). DaVinci Resolve enabled the addition of captions, images, and textual annotations and allowed precise control of annotation timing and alignment. Participants accessed them on mobile devices to answer comprehension questions, summarize content, and provide feedback. All students were enrolled in the same EFL course with two weekly sessions (100 minutes), including one 50-minute vlog-based listening task. The captioned group watched two short conversations and two short talks with captions each week, whereas the annotated group viewed the same content with textual and pictorial annotations, which are more effective than single-mode annotations (Shahipanah et al., 2025). Target words in the annotated vlogs appeared in bold in red frames with Chinese (L1) definitions, English definitions, and images. The control group was exposed to the same short conversations and short talks in audio files as those in the experimental groups, absent only the multimodal enhancements and vlog form.

To ensure that the target vocabulary was unfamiliar, a pilot study was conducted with 10 students of similar English proficiency who did not participate in the formal study. After completing the eight short conversations and eight short talks, they reviewed the scripts and underlined unfamiliar words. Based on self-reports indicating whether they knew, were unsure of, or did not know each word's meaning, 42 unfamiliar words were identified from the short conversations and 43 from the short talks. Words were classified as unfamiliar when participants reported not knowing their meaning, and only those items for which 100% of the pilot participants indicated unfamiliarity were retained in the target list, while any words failing to meet this criterion were excluded. Two English-teaching experts reviewed these words for A2 CEFR suitability, removing items that were too simple, unsuitable for annotation, or repeated. Following expert review, two English teachers independently evaluated the target vocabulary for difficulty and appropriateness. Interrater agreement was 90.5% for short conversation items (38/42) and 90.7% for short talk items (39/43); disagreements were resolved through discussion. Finally, 28 target words were selected from the short conversations and 28 from the short talks. Word frequency was analyzed using Range program (Nation & Heatley, 2002) and Compleat Web VP v.2.2 with the BNC-COCA lists (Nation, 2017). The word-family unit from the BNC-COCA list was adopted because it is widely used in lexical profiling and vocabulary research to estimate lexical coverage and text difficulty. Employing this unit allows the present study to remain comparable with previous studies that have applied the same framework. Although the two sets differed slightly, over 78% of the words fell within

the 2K–3K frequency bands, indicating comparable difficulty. Appendix C presents the two sets of words with their word frequency results.

Figure 1

Screenshot of a Captioned Vlog

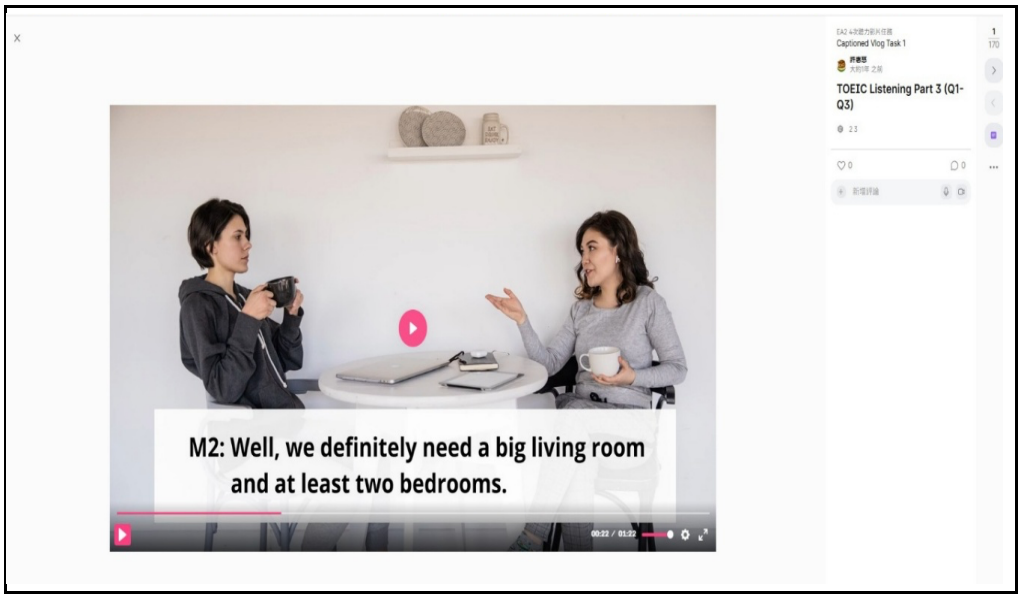


Figure 2

Screenshot of Four Captioned Vlog Tasks

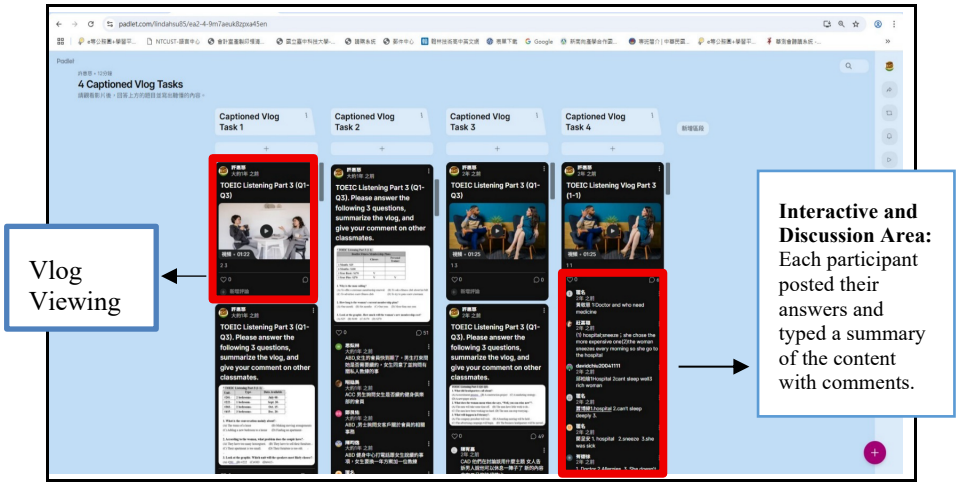


Figure 3

Screenshot of an Annotated Vlog

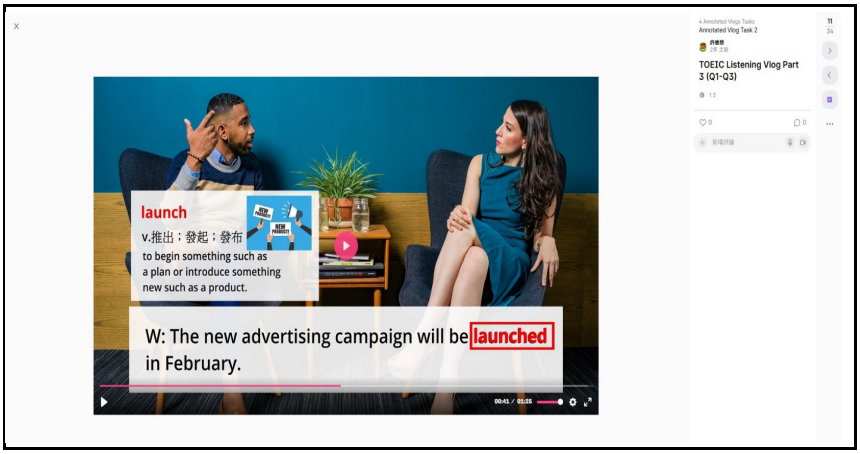
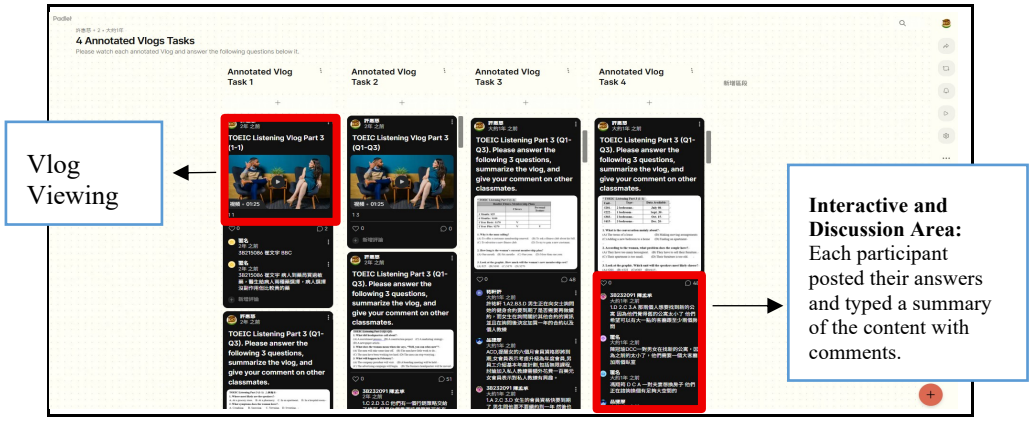


Figure 4

Screenshot of Four Annotated Vlog Tasks



Annotation Design

The design of the captioned and annotated vlog conditions was informed by principles from cognitive load theory and multimedia learning (Mayer & Moreno, 2003). As shown in Appendix B, captioned vlogs included on-screen captions of spoken dialogue to reinforce auditory input and support comprehension. Annotated vlogs incorporated additional features such as visual highlighting of target words, bilingual glosses with images, synchronized timing with spoken input, adjacent placement to the relevant dialogue, and brief on-screen display to guide attention to target vocabulary, facilitate form-meaning integration, reduce split attention, and manage cognitive load, thereby supporting more efficient lexical processing.

Research Instruments

The instruments included listening and vocabulary pre- and post-tests, a self-efficacy scale, an annotated vlog learning perception scale, and semi-structured interview questions. Except for the listening and vocabulary tests, all measures were presented in Chinese (L1) to minimize language-related misunderstanding.

Listening Pre- and Post-tests

The listening test showed acceptable internal consistency (Cronbach's $\alpha = .77$), exceeding the .70 threshold (Nunnally & Bernstein, 1994). The tests were adapted from the TOEIC developed by

Educational Testing Service, aligned with the CEFR and recognized for reliability and validity. The TOEIC listening section includes 100 items: photographs ($N = 6$), question-response ($N = 25$), conversations ($N = 39$), and talks ($N = 30$), and takes about 45 minutes. Comparable but nonidentical TOEIC-aligned tests were used as the pre-test and post-test to reduce practice effects. Materials were adapted from a TOEIC practice test book (Educational Testing Service, 2023) for instructional use. Although the official TOEIC Listening test uses a 5-495 scale, the adapted tests were scored on a 0–100 scale for classroom interpretation, and participants were not informed in advance about the post-test.

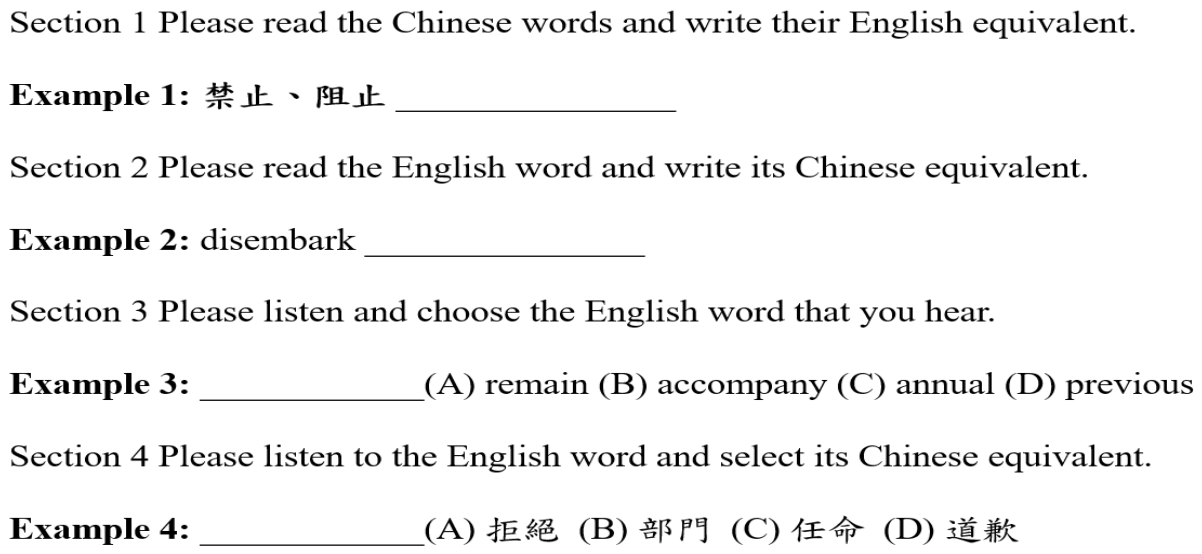
Vocabulary Pre- and Post-tests

The vocabulary test showed good internal consistency (Cronbach's $\alpha = .81$), exceeding the .80 threshold (Nunnally & Bernstein, 1994). Pre- and post-tests were developed from target vocabulary (Çakmak & Erçetin, 2018) drawn from eight short conversations and eight short talks adapted from the *TOEIC Upgrade* practice book. A pilot study and expert review identified 56 A2-level target words (28 per task type), from which 40 were randomly selected to measure incidental vocabulary retention. The tests assessed receptive and productive vocabulary gains (Laufer & Goldstein, 2004) and were administered online via Google Forms with timed, nonrevisitable sections. Each word was played once, and students provided written responses in sequential sections (productive, then receptive; see Figure 5).

Each test included two productive and two receptive sections with 10 items each, totaling 40 items in both the vocabulary pre- and post-tests. The vocabulary test comprised four sections assessing knowledge of target vocabulary across different modalities. Each section included a different subset of 10 items. Section 1 required translating Chinese words into English, assessing retrieval of English forms from L1 meanings. Section 2 required translating English words into Chinese, measuring form-meaning knowledge. Section 3 required selecting the English word heard from multiple options, assessing auditory recognition of lexical forms. Section 4 required selecting the Chinese equivalent of a spoken English word, measuring mapping of spoken forms to meaning. Different subsets of items were used across sections to reduce fatigue and practice effects while assessing vocabulary knowledge in written and aural modalities. Each correct answer was scored as 2.5 points, for a maximum score of 100, with dichotomous scoring (correct versus incorrect) and no partial credit. The post-test used the same target items as the pre-test to measure item-specific learning and retention, but the order was rearranged to reduce recall effects, and participants were not informed in advance to minimize test-retest effects.

Figure 5

Representative Examples of Vocabulary Test Items



Self-efficacy

The self-efficacy scale demonstrated excellent internal consistency (Cronbach's α = .94). Reliability values above .70 are acceptable, above .80 are good, and above .90 are excellent for research use (Nunnally & Bernstein, 1994). The scale was adapted from the self-efficacy section of the motivated strategies for learning survey developed by Pintrich et al. (1991). It comprised eight items on a 7-point Likert scale and was administered before and after the intervention to measure changes in learner self-efficacy (see Appendix D).

Annotated Vlog Learning Perception Survey

The annotated vlog learning perception scale showed excellent reliability (α = .92). Reliability coefficients of .70 are acceptable, .80 are good, and .90 are excellent (Nunnally & Bernstein, 1994). Adapted from Aldukhayel's (2021a) vlog learning perception scale, it includes two sections: (1) effectiveness of annotated vlogs in enhancing English learning (six items on a seven-point Likert scale) and (2) integration of annotated vlogs into English listening learning (four items on a seven-point Likert scale) (see Appendix D).

Semi-structured Interview

A semi-structured interview was conducted to explore participants' perceptions of the vlog-based learning. Five participants from the captioned vlog group and five from the annotated vlog group were randomly selected for a 20-minute interview. All interviews were conducted by the same researcher, with participants responding in Chinese (their L1), and were audio-recorded. Random sampling was used to reduce selection bias (Shadish et al., 2002). In Appendix E, four open-ended questions on task helpfulness, differences from previous listening tasks, and willingness to persist were adapted from Aldukhayel (2021a).

Data Collection

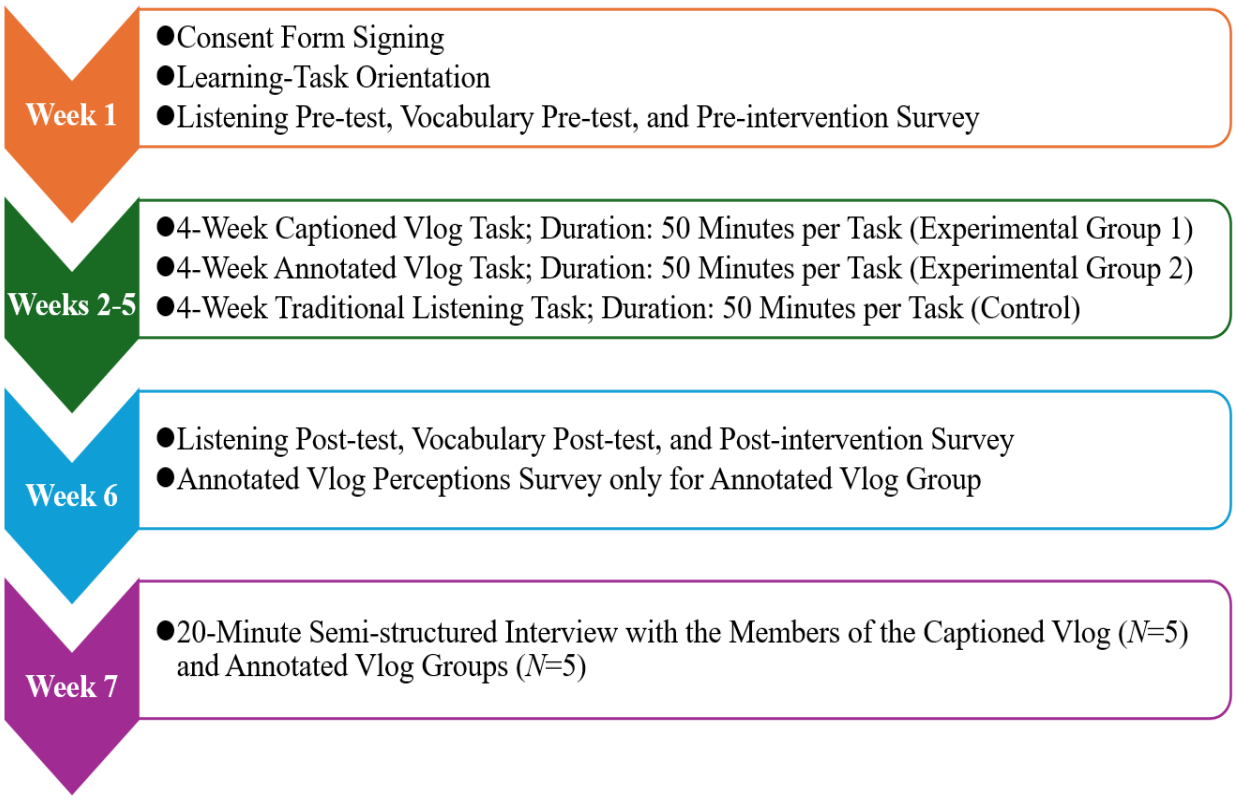
Data collection was conducted in four stages in 2023 (see Figure 6). First, procedures were explained, consent was obtained, and participants completed the listening and vocabulary pre-tests and pre-intervention survey. Second, all groups received a four-week intervention with weekly 50-minute sessions (15 minutes listening, 35 minutes follow-up). The vlog groups used captioned or annotated vlogs with Padlet-based questions, discussions, and peer feedback, whereas the control group completed instructor-led tasks; time on task was equivalent. Third, participants completed the post-tests and post-intervention survey, and the annotated vlog group completed a learning perceptions survey. Finally, five students from each experimental group were randomly selected for 20-minute semi-structured interviews.

Data Analysis

An analysis of covariance (ANCOVA) was used to examine the effects of annotated vlog learning on listening, vocabulary, and self-efficacy, with group as the independent variable, post-test scores as dependent variables, and pre-test scores as covariates to control for initial differences. Descriptive statistics and Cronbach's alpha assessed the reliability of the learning perceptions survey administered to the annotated vlog group. Semi-structured interview transcripts were analyzed using Wolcott's (1994) framework of description, analysis, and interpretation. Transcripts were verified, inductively coded for recurring themes, and interpreted in relation to the research questions. For example, comparisons of vlog-based and audio-only listening were coded as enhanced clarity and improved comprehension, indicating that multimodal features, particularly visual and annotated support, facilitated listening comprehension more effectively than audio-only instruction. Based on the coding scheme, two experienced English instructors independently coded the data across five themes. The coders agreed on 38 of the 40 units (95.0%), indicating high intercoder reliability (Cohen's κ = .93). Disagreements were resolved through discussion until consensus was reached, and the final coding reflected the agreed classifications.

Figure 6

Data Collection Procedure



Results

Vocabulary Retention and Listening Test Results

Table 1 shows pre- to post-test improvement in vocabulary and listening across all groups. Vocabulary scores increased from 37.91 to 43.95 (control group), 42.27 to 49.66 (captioned vlog group), and 41.93 to 52.05 (annotated vlog group). Listening scores increased from 29.35 to 31.77 (control group), 33.23 to 44.68 (captioned vlog group), and 31.68 to 50.14 (annotated vlog group). The annotated vlog group obtained the highest means in both vocabulary retention and listening comprehension.

Table 1

Descriptive Statistics of Vocabulary Retention and Listening Test Results

Item	Group	*N*	*M*		*SD*	
			Pre-test	Post-test	Pre-test	Post-test
Vocabulary	Control	43	37.91	43.95	5.694	6.129
	Captioned Vlog	44	42.27	49.66	6.048	8.380
	Annotated Vlog	44	41.93	52.05	8.506	10.191
	Mean		40.73	48.59	7.102	9.014
Listening	Control	43	29.35	31.77	7.606	6.651
	Captioned Vlog	44	33.23	44.68	8.978	12.374
	Annotated Vlog	44	31.68	50.14	8.594	12.489
	Mean		31.44	42.27	8.504	13.264

Because random sampling and assignment were not fully feasible, this study used a quasi-experimental design with statistical control for confounding variables. ANCOVA typically treats pre-test scores as covariates and post-test scores as dependent variables and requires independence and homogeneity of variance. Before conducting ANCOVA, preliminary tests were performed to examine the homogeneity of pretest variances and covariate balance across the three classes. Levene's tests indicated unequal variances in vocabulary pretest scores, $F(2, 128) = 5.173$, $p = .007$ (< .05), whereas no significant variance differences were found in listening pretest scores, $F(2, 128) = 0.106$, $p = .899$ (> .05). A one-way ANOVA showed significant between-class differences in vocabulary pretest scores, $F(2, 128) = 5.409$, $p = .006$ (< .05), indicating baseline imbalance across the classes. No significant between-class differences were found in listening pretest scores, $F(2, 128) = 2.336$, $p = .101$ (> .05). Accordingly, the vocabulary outcome was analyzed using quasi-ANCOVA with a group-mean-centered covariate following Huitema (2011), whereas the listening outcome was analyzed using ANCOVA.

Quasi-ANCOVA involved mean-centering covariates within groups, testing regression homogeneity, and conducting the analysis with centered covariates. For vocabulary, the vocabulary pretest scores were group-mean-centered before testing the homogeneity of regression slopes. The interaction between class and the centered vocabulary pretest scores was not significant, $F(2, 125) = 1.226$, $p = .297$ (> .05). ANCOVA was used for listening, with the homogeneity of regression slopes tested before the analysis. The interaction between class and the listening pretest scores was not significant, $F(2, 125) = 1.951$, $p = .146$ (> .05), indicating no significant differences in regression slopes among the three classes. Thus, the homogeneity of regression slopes assumption was met for both vocabulary and listening (see Table 2).

Table 2

Results of Homogeneity Tests of Regression Coefficients

	SV	***SS***	***df***	***MS***	***F***	***p***
Vocabulary	RC	132.791	2	66.396	1.226	0.297
	Error	6767.896	125	54.143		
	Sum	6900.687	127			
Listening	RC	418.743	2	209.371	1.951	0.146
	Error	13410.982	125	107.288		
	Sum	13829.725	127			

Note. SV = Source of Variance. *SS* = Sum of Squares. *df* = Degrees of Freedom. *MS* = Mean Square. *F* = *F* Statistic. *p* = Probability Value. RC = Regression Coefficients.

In the quasi-ANCOVA, Levene's tests indicated nonsignificant results for vocabulary, $F(2, 128) = 1.684$, $p = .190$ (>.05), confirming homogeneity of variance across groups. The ANCOVA was conducted for listening. Levene's test was significant, $F(2, 128) = 7.427$, $p < .001$, indicating that the homogeneity of error variances assumption was not met. Therefore, HC3 heteroskedasticity-consistent standard errors were used for robust inference. After controlling for the listening pretest and using the control group as the reference group, HC3-adjusted results showed that the captioned vlog group scored significantly higher than the control group on the listening posttest, $B = 11.436$, $SE_{HC3} = 2.028$, $t = 5.638$, $p < .001$. The annotated vlog group also scored significantly higher than the control group, $B = 17.479$, $SE_{HC3} = 2.157$, $t = 8.103$, $p < .001$. These differences remained significant after accounting for heteroskedasticity, supporting the robustness of the findings (Hayes & Cai, 2007; Long & Ervin, 2000). In Table 3, tests of between-subject effects showed significant adjusted mean differences in vocabulary ($p = .000$) and listening ($p = .000$). Partial η^2 indicates the proportion of variance explained by a factor, with $0.01 \leq \eta^2 < 0.06$ representing a small effect, $0.06 \leq \eta^2 < 0.14$ a medium effect, and $\eta^2 \geq 0.14$ a large effect (Cohen, 2013). The effect size for vocabulary was large (partial $\eta^2 = .179$, statistical power = .998) and the effect size for listening was also large (partial $\eta^2 = .327$, statistical power = 1.000).

Table 3

Results of Analysis of Covariance on the Listening and Vocabulary Post-tests

	SV	*SS*	*df*	*MS*	*F*	*p*	*partial η2*	*1-β*
Vocabulary	BG	1500.038	2	750.019	13.803	0.000	0.179	0.998
	WG	6900.687	127	54.336				
	Total	8400.725	129					
Listening	BG	6714.531	2	3357.265	30.830 [a]	0.000	0.327	1.000
	WG	13829.725	127	108.895				
	Total	20544.256	129					

Note. [a] is based on HC3 heteroskedasticity-robust standard errors. *SV* = Source of Variance. *SS* = Sum of Squares. *df* = Degrees of Freedom. *MS* = Mean Square. BG = Between Groups. WG = Within Groups.

Overall, post-test vocabulary and listening means were adjusted for baseline scores so that group differences reflected intervention effects rather than pre-existing disparities. Quasi-ANCOVA showed significant group differences in vocabulary, $F(2,127) = 13.803$. After adjusting for the vocabulary pre-test, the annotated vlog group (52.045) scored significantly higher than the captioned vlog group (49.659) and the control group (43.953). ANCOVA revealed significant differences were also found in listening, $F(2,127) = 30.830$; after adjusting for the listening pre-test, the annotated vlog group (50.042) scored significantly higher than the captioned vlog group (43.999) and the control group (32.563; Tables 3 and 4).

Table 4

Summary of Adjusted Means and Standard Deviations

	Group	***N***	***M***	***SD***	**95% confidence interval**	
					L	**U**
Vocabulary	Control	43	43.953[a]	1.122	41.733	46.174
	Captioned Vlog	44	49.659[a]	1.109	47.464	51.855
	Annotated Vlog	44	52.045[a]	1.109	49.850	54.241
Listening	Control	43	32.563	1.608	29.382	35.744
	Captioned Vlog	44	43.999	1.585	40.861	47.136
	Annotated Vlog	44	50.042	1.573	46.929	53.156

Note. [a]indicates the covariates presented in the model were estimated using the following value: Centralization_pre-test(vocabulary) = .0000.

Self-Efficacy

Table 5 shows pre- to post-test improvement in self-efficacy across all groups. Scores increased from 3.017 to 3.398 in the control group, 4.119 to 5.185 in the captioned vlog group, and 5.477 to 5.5 in the annotated vlog group, with greater variability observed at the post-test.

Table 5

Descriptive Statistics of Self-Efficacy

Item	**Group**	***N***	***M***		***SD***	
			Pre-test	**Post-test**	**Pre-test**	**Post-test**
Self-efficacy	Control	43	3.017	3.398	0.545	0.661
	Captioned Vlog	44	4.119	5.185	0.281	0.844
	Annotated Vlog	44	5.477	5.500	0.712	0.762
Mean			4.204	4.694	0.513	0.756

Under the assumptions of independence and homogeneity of variance, Levene's test for self-efficacy was significant, Levene = 13.788, $p = .000$ (< .05), indicating unequal variance across the three groups. One-way ANOVA also revealed a significant difference in self-efficacy, $F(2,128) = 224.31$, $p = .000$ (< .05), violating the independence assumption and indicating that standard ANCOVA was inappropriate. Therefore, following Huitema (2011), quasi-ANCOVA with mean-centered covariates was applied.

Quasi-ANCOVA was conducted in three steps: mean-centering the covariate, testing the homogeneity of regression coefficients, and analyzing the results with the centered covariate. After centering the self-efficacy pre-test scores, the regression coefficients across the three groups were equivalent, $F(2,125) = 0.212$, $p = .809$ (>.05), satisfying the homogeneity assumption (see Table 6). Therefore, the data were suitable for quasi-ANCOVA.

Table 6

Results of Homogeneity Test of Regression Coefficients

	SV	***SS***	***df***	***MS***	***F***	***p***
Self-efficacy	Regression Coefficients	0.232	2	0.116	0.212	0.809
	Error	68.306	125	0.495		
	Sum	68.538	127			

Note. SV = Source of Variance. *SS* = Sum of Squares. *df* = Degrees of Freedom. *MS* = Mean Square. *F* = *F* Statistic. *p* = Probability Value.

In the quasi-ANCOVA, Levene's test for self-efficacy was nonsignificant, $F(2,128) = 2.01$, $p = .143$ (>.05), indicating homogeneity across the three groups. Partial η^2 represents the proportion of variance explained by a factor after controlling for other variables, with $0.01 \leq \eta^2 < 0.06$ indicating a small effect, $0.06 \leq \eta^2 < 0.14$ a medium effect, and $\eta^2 \geq 0.14$ a large effect (Cohen, 2013). In Table 7, the between-subject effects test showed a significant group effect ($p = .000 < .05$). The effect size for self-efficacy was large (partial $\eta^2 = .62$, statistical power = 1).

Table 7

Results of Analysis of Covariance of Self-Efficacy

	SV	***SS***	***df***	***MS***	***F***	***p***	**partial η2**	**1-β**
Self-efficacy	Between Groups	111.358	2	55.679	101.892	0.000	.62	1
	Within Groups	68.306	125	.546				
	Total	185.35	127					

Note. SV = Source of Variance. *SS* = Sum of Squares. *df* = Degrees of Freedom. *MS* = Mean Square. *F* = *F* Statistic. *p* = Probability Value.

As shown in Table 7, quasi-ANCOVA revealed significant differences among the three groups in self-efficacy, $F(2,125) = 101.892$. Post-intervention means were adjusted for baseline self-efficacy scores so that group differences reflected intervention effects rather than pre-existing disparities. After adjustment, the annotated vlog group (5.5) scored significantly higher than the captioned vlog group (5.185) and the control group (3.398; Table 8).

Table 8

Summary of Adjusted Means and Standard Deviations

	Group	*M*	*SD*	95% Confidence Interval	
				L	U
Self-efficacy	Control	3.398[a]	0.113	3.175	3.621
	Captioned Vlog	5.185[a]	0.111	4.964	5.405
	Annotated Vlog	5.500[a]	0.111	5.279	5.721

Note. [a]indicates the covariates presented in the model were estimated using the following value: Centralization_pre-test(vocabulary) = .0000.

Perceptions of Learning Using Annotated Vlogs

In the annotated vlog group, all participants believed that annotated vlogs benefitted their English learning (*M* = 5.46; *SD* = 1.139). Additionally, all participants responded positively to the use of annotated vlogs in English listening learning (*M* =5.61; *SD* = 1.163; Table 9).

Table 9

Results of Descriptive Analysis of Annotated Vlog Learning Perception Survey

	Item	*M*	*SD*
Effectiveness of Annotated Vlog on English Learning	1	5.55	0.975
	2	5.61	0.970
	3	5.41	1.317
	4	6.00	1.100
	5	5.00	1.347
	6	5.18	1.126
	Average	5.46	1.139
Integration of Annotated Vlog into English Listening Learning	7	5.34	1.397
	8	5.64	1.163
	9	5.73	1.169
	10	5.73	0.924
	Average	5.61	1.163
Overall Average		5.519	1.149

Qualitative Analysis of Interview Data from the Experimental Groups

The qualitative analysis followed Wolcott's (1994) framework. Themes were identified inductively from recurring patterns in the interview transcripts and grouped into broader categories when they appeared consistently across participants and were relevant to the research questions. Five overarching themes were identified in the interview data, as summarized in Table 10. Three were shared across both groups: vlog-based listening was engaging, the vlog interface was interactive, and repeated listening improved comprehension. The captioned vlog group reported that although captions helped them follow the audio, unfamiliar lexical items disrupted comprehension, whereas the annotated vlog group indicated that textual and pictorial annotations supported both listening and vocabulary learning.

The qualitative themes corroborate the quantitative findings. Participants in both vlog groups regarded vlog-based listening as clearer, more engaging, and more interactive than audio-only instruction, and noted that repeated listening improved comprehension. This aligns with survey support for annotated vlogs and with post-test gains over pre-test performance in listening comprehension and vocabulary retention. Additionally, group differences further explain these quantitative outcomes. Learners in the captioned vlog group reported difficulty with unfamiliar vocabulary, which may have limited comprehension despite captions. By contrast, those in the annotated vlog group indicated that textual and pictorial annotations supported both comprehension and vocabulary learning, accounting for their

stronger performance compared with the captioned group.

Table 10

Emergent Themes in Students' Perception of Annotated Vlogs for Listening Learning

Group	Theme (Frequency)	Representative Responses
Annotated and Captioned	Engaging vlog-based delivery of listening content (10 of 10)	*"Vlog-assisted listening was more interesting and engaging than audio-only exercises."* *"Compared with audio-only listening, annotated vlogs were more engaging and easier to understand."* *"Combining viewing and listening in vlog-based practice increased engagement and helped learners focus on the content."*
	Interactive vlog interface (8 of 10)	*"Discussing listening content and answers in the comments improved comprehension and engagement."* *"Reading classmates' answers and summaries clarified previously misunderstood content."* *"Sharing answers and giving likes in the comments increased my motivation to practice listening."*
	Improved comprehension through repeated listening (7 of 10)	*"Replaying the vlog helped me become more familiar with the listening passage."* *"Repeated viewing of annotated vlogs improved listening speed and comprehension."* *"Repeated listening with annotated captions improved listening comprehension."*
Captioned	Unfamiliar vocabulary in captions hindering listening comprehension (3 of 5)	*"Unfamiliar key words in the captioned vlog made the passage difficult to understand."* *"Although captions helped me follow the listening, unfamiliar vocabulary sometimes disrupted comprehension."* *"Unfamiliar vocabulary in the captions affected listening comprehension."*
Annotated	Textual and pictorial annotations facilitating listening and vocabulary learning (4 of 5)	*"Annotated keywords with textual and pictorial cues helped me notice and understand new words while listening."* *"Keyword annotations with Chinese and English definitions and images clarified key information and supported vocabulary learning."* *"Annotated vlogs improved listening comprehension and vocabulary learning through simultaneous viewing and listening."*

Discussion

This study examined the effects of annotated vlogs delivered through mobile devices on Taiwanese EFL learners' listening comprehension and vocabulary retention. Research Questions 1 and 2 addressed the effects of captioned and annotated vlogs on these outcomes. Pre- and post-test results showed significant differences among the three instructional conditions, with the annotated vlog group outperforming the captioned vlog and control groups in both listening comprehension and vocabulary retention. Although the captioned vlog group outperformed the control group, suggesting that captions facilitated form-meaning connections during video viewing (Vanderplank, 2016), annotations provided more effective

support through concise lexical and semantic information via textual and pictorial cues. Unlike captions, which may shift attention toward reading and limit auditory processing (Winke et al., 2010), annotations allowed learners to maintain focus on spoken input while accessing meaning support. These findings align with studies demonstrating the benefits of multimodal annotations for listening comprehension and vocabulary learning (Çakmak & Erçetin, 2018; Jones & Plass, 2002), although some research suggests annotations may increase cognitive load (Plass, 2003). Overall, the results indicate that vlog annotations can effectively enhance listening comprehension and vocabulary retention, highlighting the importance of instructional design in multimedia-assisted language learning (Mayer, 2020).

The performance of the captioned vlog group also merits attention. Although they performed less well than the annotated vlog group, captions likely supported listening comprehension by reinforcing spoken input with written text, allowing learners to confirm what they heard. However, because captions provide less explanatory support than annotation-enhanced vlogs, their effects on vocabulary retention and deeper comprehension were weaker, and unfamiliar lexical items sometimes disrupted comprehension. Nevertheless, captioned vlogs may still serve as a useful scaffold for learners who require additional support in processing spoken input (Aldukhayel, 2021b; Yeldham, 2018).

Research Question 3 examined the influence of captioned and annotated vlogs on learners' self-efficacy. The results showed that the annotated vlog group demonstrated significantly greater self-efficacy than the other groups after the intervention, suggesting that annotations and multimedia support helped learners cope with listening challenges and increased engagement and persistence. In addition, both vlog groups showed greater gains in self-efficacy than the control group. This finding aligns with studies indicating that digital technologies such as vlogs can enhance self-efficacy and promote active participation and improved learning outcomes (Major, 2026; Zadorozhnyy & Lee, 2025). The improvement in self-efficacy may also reflect increased confidence and a sense of achievement during the learning process.

Research Question 4 examined learners' perceptions of the learning process when engaging in listening tasks with annotated vlogs. Learning perceptions enable learners to evaluate strategies and adjust learning behaviors (Zhao et al., 2025). The perceptions of learning using annotated vlogs scale indicated favorable attitudes toward their use for English listening and vocabulary learning. Annotations and multimedia support helped learners cope with listening challenges and increased engagement and persistence. Semi-structured interviews in both experimental groups (captioned and annotated) also showed that the engaging vlog delivery, interactive interface, and opportunities for repeated practice enhanced learners' focus. Vlogs allowed learners to adjust pace and strategies, while repeated viewing reinforced retention and increased confidence and autonomy. Most learners reported that textual and pictorial annotations supported listening comprehension and vocabulary retention, particularly when captions alone were insufficient for unfamiliar words, consistent with previous studies (Jones & Plass, 2002; Shahipanah et al., 2025). Overall, the findings indicate that annotated vlogs provided stronger learning support than captioned vlogs.

Pedagogical Implications

This study has three pedagogical implications. First, in comparable instructional settings, annotated vlogs can be effectively integrated into listening instruction by embedding task-oriented activities that enable instructors to monitor participation, provide timely feedback, and foster instructor-student interaction. Interactive features such as commenting facilitate error correction, strengthen comprehension, and promote peer discussion, creating a more engaging and learner-centered environment than that created during traditional audio-based instruction. However, producing annotated vlogs requires substantial time and technical support, which may limit scalability; reusable annotation templates may help address this issue. Second, annotated vlogs encourage active engagement by allowing learners to use peers' comments to clarify listening comprehension. As one participant noted, "Reading peers' responses and the vlog summaries in the comment section allowed me to correct my understanding of parts of the vlog that I initially misunderstood or could not follow." These interactions help learners verify and refine their comprehension. Finally, mobile delivery allows students to complete listening tasks at their own pace,

freeing class time for interaction and feedback. It supports individualized learning and collaborative engagement, although small screens and frequent notifications may disrupt attention. Teachers can mitigate this limitation by using shorter video segments and asking students to disable notifications during tasks.

Limitations and Directions for Future Research

This study has seven limitations. First, annotated vlogs were delivered only through Padlet; other platforms may produce different outcomes. Future studies should compare mobile-delivered annotated vlogs with those on platforms such as YouTube in terms of listening comprehension and vocabulary retention. Second, all participants had the same English proficiency and were from a single university; future research should include learners from multiple institutions and proficiency levels to enhance generalizability. Third, only one psychological variable, self-efficacy, was examined. Future research should also consider enjoyment, motivation, and learning anxiety to better understand individual differences. Fourth, the absence of comparative analysis of learning perceptions across the three groups limited understanding of group differences; a survey to examine perceptual differences among groups could address this issue in future studies. Fifth, lexical profiling used the BNC-COCA word family lists (Nation, 2017). Although widely used, this framework assumes that learners who know a base word can recognize its inflected and derived forms. However, learners' knowledge of derivational morphology may be incomplete, meaning word-family analyses may underestimate lexical difficulty, especially for intermediate learners in East Asian EFL contexts (Milliner et al., 2024). Future research could examine lexical coverage using alternative units (e.g., lemmas or word types) and consider learners' knowledge of derivational morphology to provide more accurate estimates of lexical difficulty. Sixth, causal interpretation is limited. Although the study compared captioned vlogs, annotated vlogs, and an auditory-only control condition to examine input modality effects on vocabulary retention and listening comprehension, the vlog-based conditions may also involve interface affordances, particularly potential interactivity. Therefore, modality effects may be confounded with platform features. Future research should examine modality and interface affordances separately to clarify their contributions to vocabulary learning and listening comprehension. Finally, the study lacked longitudinal follow-up; future research should include a delayed post-test to examine long-term vocabulary retention.

Conclusion

For the participants involved in the current study, annotated vlogs appeared to support listening comprehension and vocabulary development. Vlogs delivered via Padlet foster instructor-student and peer interactions through real-time feedback. Additionally, mobile delivery may enable the instructor to flexibly monitor learners' progress and provide timely support, while also allowing participants to learn at their own pace during class. By transcending the rigidity of conventional video-based instruction, mobile-delivered annotated vlogs can provide learners in educational settings similar to that reported in this study pedagogical innovation that enhances engagement and learning outcomes.

Funding

This study was funded by grants to the author by the National Science and Technology Council, Taiwan, R.O.C. (NSTC 112－2410－H－167－008－).

References

Aldera, A. S., & Mohsen, M. A. (2013). Annotations in captioned animation: Effects on vocabulary learning and listening skills. *Computers & Education, 68*, 60–75. http://dx.doi.org/10.1016/j.compedu.2013.04.018

Aldukhayel, D. (2021a). Vlogs in L2 listening: EFL learners' and teachers' perceptions. *Computer Assisted Language Learning, 34*(8), 1085–1104. https://doi.org/10.1080/09588221.2019.1658608

Aldukhayel, D. (2021b). The effects of captions on L2 learners' comprehension of Vlogs. *Language Learning & Technology, 25*(2), 178–191. https://doi.org/10.64152/10125/73439

Anderson, P. (2017). *TOEIC upgrade.* Compass Publishing.

Arndt, H., & Woore, R. (2018). Vocabulary learning from watching YouTube videos and reading blog posts. *Language Learning and Technology, 22*(3), 124–142. https://doi.org/10.64152/10125/44660

Bai, B., Chao, G. C. N., & Wang, C. (2019). The relationship between social support, self-efficacy, and English language learning achievement in Hong Kong. *TESOL Quarterly, 53*(1), 208–221. https://doi.org/10.1002/tesq.439

Bandura, A., Freeman, W. H., & Lightsey, R. (1997). Self-efficacy: The exercise of control. *Journal of Cognitive Psychotherapy, 13*, 158–166. https://doi.org/10.1891/0889-8391.13.2.158

Bandura, A., & Schunk, D.H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology, 41*(3), 586–598. https://psycnet.apa.org/doi/10.1037/0022-3514.41.3.586

Çakmak, F., & Erçetin, G. (2018). Effects of gloss type on text recall and incidental vocabulary learning in mobile-assisted L2 listening. *ReCALL, 30*(1), 24-47. https://doi.org/10.1017/S0958344017000155

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293–332. https://doi.org/10.1207/s1532690xci0804_2

Chao, Y. G. (2013). Factors affecting college EFL learners' listening comprehension and listening problems. *NCUE Journal of Humanities, 8*, 71–84.

Chen, C. M., & Chen, I. C. (2021). The effects of video-annotated listening review mechanism on promoting EFL listening comprehension. *Interactive Learning Environments, 29*(1), 83–97. https://doi.org/10.1109/IIAI-AAI.2018.00037

Chen, C. M., Li, M. C., & Lin, M. F. (2022). The effects of video-annotated learning and reviewing system with vocabulary learning mechanism on English listening comprehension and technology acceptance. *Computer Assisted Language Learning, 35*(7), 1557–1593. https://doi.org/10.1080/09588221.2020.1825093

Chen, M. C., & Lin, H. J. (2009). Self-efficacy, foreign language anxiety as predictors of academic performance among professional program students in a general English proficiency writing test. *Perceptual & Motor Skills, 109*(2), 420–430. https://doi.org/10.2466/pms.109.2.420-430

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge. https://doi.org/10.4324/9780203771587

Educational Testing Service. (2023). *TOEIC® listening and reading test official test-preparation guide* (Vol. 8). Chunshin Publishing.

Hao, Y., Lee, K. S., Chen, S. T., & Sim S. C. (2019). An evaluative study of a mobile application for middle school students struggling with English vocabulary learning. *Computers in Human Behavior, 95*, 208–216. https://doi.org/10.1016/j.chb.2018.10.013

Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods, 39*(4), 709-722. https://doi.org/10.3758/BF03192961

He, J., Gao, B., & Wang, Y. (2025). From attraction to retention: How vlogger attributes, vlog news content quality, and platform features affect continuance intention of vlog news. *International Journal of Human–Computer Interaction, 41*(14), 9151–9172. https://doi.org/10.1080/10447318.2024.2423334

Hou, J., Lee, J. F., & Doherty, S. (2024). The effects of mobile-assisted reading on incidental L2 word learning: a processing perspective. *Computer Assisted Language Learning, 37*(7), 2254–2289. https://doi.org/10.1080/09588221.2022.2157445

Hsieh, Y. (2020). Effects of video captioning on EFL vocabulary learning and listening comprehension. *Computer Assisted Language Learning, 33*(5–6), 567–589. https://doi.org/10.1080/09588221.2019.1577898

Hsu, C. K. (2015). Learning motivation and adaptive video caption filtering for EFL learners using handheld devices. *ReCALL, 27*(1), 84–103. https://doi.org/10.1017/S0958344014000214

Hsu, C. K., Hwang, G.-J., Chang, Y.-T., & Chang, C.-K. (2013). Effects of video caption modes on English listening comprehension and vocabulary acquisition using handheld devices. *Journal of Educational Technology & Society, 16*(1), 403–414. https://www.jstor.org/stable/jeductechsoci.16.1.403

Hsu, H. T. (2018). Incidental professional vocabulary acquisition of EFL business learners: Effect of captioned video with glosses. *The JALT CALL Journal, 14*(2), 119–142. https://doi.org/10.29140/jaltcall.v14n2.j227

Huang, H. W. (2021). Effects of smartphone-based collaborative vlog projects on EFL learners' speaking performance and learning engagement. *Australasian Journal of Educational Technology, 37*(6), 18–40. https://doi.org/10.14742/ajet.6623

Huitema, B. (2011). *The analysis of covariance and alternatives: Statistical methods for experiments, quasi-experiments, and single-case studies*. Wiley. https://doi.org/10.1002/9781118067475

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson, *Cognition and second language instruction* (pp. 258–286). Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.011

Jones, L. & J. Plass (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal, 86*, 546–561. https://doi.org/10.1111/1540-4781.00160

Kay, R. H. (2012). Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior, 28*(3), 820–831. https://doi.org/10.1016/j.chb.2012.01.011

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language learning, 54*(3), 399–436. https://doi.org/10.1111/j.0023-8333.2004.00260.x

Li, R. (2024). Effects of mobile-assisted language learning on foreign language learners' speaking skill development. *Language Learning & Technology, 28*(1), 1–26. https://doi.org/10.64152/10125/73553

Lin, C. C., & Yu, Y. C. (2017). Effects of presentation modes on mobile-assisted vocabulary learning and cognitive load. *Interactive Learning Environments, 25*(4), 528–542. https://doi.org/10.1080/10494820.2016.1155160

Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL, 31*(3), 293–311. https://doi.org/10.1017/S0958344019000065

Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician, 54*(3), 217–224. https://doi.org/10.1080/00031305.2000.10474549

Major, L. (2026). Using vlogs as asynchronous peer teaching to enhance student autonomy. *Technology, Pedagogy and Education*, *35*(2), 297–313. https://doi.org/10.1080/1475939X.2025.2531877

Matthews, J., Masrai, A., Lange, K., McLean, S., Alghamdi, E. A., Kim, Y. A., Shinhara, Y., & Tada, S. (2024). Exploring links between aural lexical knowledge and L2 listening in Arabic and Japanese speakers: A close replication of Cheng, Matthews, Lange and McLean (2022). *TESOL Quarterly, 58*(1), 63–89. https://doi.org/10.1002/tesq.3212

Mayer, R. E. (2020). *Multimedia learning* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781316941355

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Merriam-Webster. (2022). *Vlogging definition & meaning.* https://www.merriam-webster.com/dictionary/vlogging#h1

Milliner, B., Lange, K., Matthews, J., & Umeki, R. (2024). Examining EFL learners' comprehension of derivational forms: The role of overlap with base word knowledge, word frequency, and contextual support. *Language Teaching Research, 28*(3), 1–26. https://doi.org/10.1177/13621688231225704

Mills, N., Pajares, F., & Herron, C. (2006). A reevaluation of the role of anxiety: Self-efficacy, anxiety, and their relation to reading and listening proficiency. *Foreign Language Annals, 39*(2), 276–295. https://doi.org/10.1111/j.1944-9720.2006.tb02266.x

Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning, 31*(1–2), 1–26. https://doi.org/10.1080/09588221.2017.1375960

Nation, I. S. P. (2017). *The BNC/COCA Level 6-word family lists* (Version 1.0.0) [Data file]. Victoria University of Wellington. http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

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

Nation, I. S. P., & Heatley, A. (2002). *Range: A program for the analysis of vocabulary in texts* [Computer software]. Victoria University of Wellington. http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

Newton, J. M., & Nation, I. S. (2020). *Teaching ESL/EFL listening and speaking*. Routledge. https://doi.org/10.4324/9780429203114

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill. https://doi.org/10.1177/014662169501900308

Okumuş Dağdeler, K. (2023). A systematic review of mobile-assisted vocabulary learning research. *Smart Learning Environments, 10*(1), Article 19. https://doi.org/10.1186/s40561-023-00235-z

Oyama, R. (2022). Exploring the relationships between enjoyment, self-efficacy, engagement, and vocabulary learning in Japanese learners of English. *Journal of Asia TEFL, 19*(4), 1163–1180. https://doi.org/10.18823/asiatefl.2022.19.4.2.1163

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).* University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior, 19*(2), 221–243. https://doi.org/10.1016/S0747-5632(02)00015-8

Ragazou, V., & Karasavvidis, I. (2023). Effects of signaling and practice types in video-based software training. *Education Sciences, 13*(6), Article 602. https://doi.org/10.3390/educsci13060602

Shadish, W. R. (2002). Revisiting field experimentation: field notes for the future. *Psychological methods, 7*(1), Article 3. https://doi.org/10.1037/1082-989x.7.1.3

Shahipanah, A., Khajavy, G. H., & Shirvan, M. E. (2025). The effect of textual and textual-pictorial glosses on incidental vocabulary learning in mobile-assisted listening. *ReCALL, 37*(1), 79–95. https://doi.org/10.1017/S0958344024000193

Sila, S. F., & Gunawan, M. H. (2025). The utilization of YouTube for extensive listening: Voices from the lecturers and students. *FRASA: English Education and Literature Journal, 6*(2), 182–200. https://doi.org/10.47701/frasa.v6i2.5040

Sung, Y. T., Chang, K. E., & Liu, T. C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education, 94*, 252–275. https://doi.org/10.1016/j.compedu.2015.11.008

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology, 14*, 50–73. https://doi.org/10.64152/10125/44214

Tabrizi, H. M., & Saeidi, M. (2015). The relationship among Iranian EFL learners' self-efficacy, autonomy and listening comprehension ability. *English Language Teaching, 8*(12), 158–169. https://doi.org/10.5539/elt.v8n12p158

Teng, F. (2022). Vocabulary learning through videos: captions, advance-organizer strategy, and their combination. *Computer Assisted Language Learning, 35*(3), 518–550. https://doi.org/10.1080/09588221.2020.1720253

Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics, 24*(3), 3–25. https://doi.org/10.1017/S0267190504000017

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action* (Vol. 12). Routledge. https://doi.org/10.4324/9780203843376

Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching*. Palgrave Macmillan. https://doi.org/10.1057/978-1-137-50045-8

Vygotsky, L. (1962). *Thought and language.* In L. Vygotsky, E. Hanfmann, & G. Vakar (Eds.), Thought and language (pp. 119–153). Boston Review.

Wang, C. H., Harrison, J., Cardullo, V., & Xi, L. (2017). Exploring the relationship among international students' English self-efficacy, using English to learn self-efficacy, and academic self-efficacy. *Journal of International Students, 8*(1), 233–250. https://doi.org/10.32674/jis.v8i1.163

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology, 14*(1), 65–86. https://doi.org/10.64152/10125/44203

Wolcott, H.F. (1994). *Transforming qualitative data: description, analysis and interpretation*. Sage.

Wood, M. (2019). *What makes a vlog a vlog?* Diggit Magazine. https://www.diggitmagazine.com/academic-papers/what-makes-Vlog-Vlog

Ye, Y., Pan, L., & Gu, Q. (2025). The effect of Chinese-related short videos on Thai students' Chinese language learning motivation and self-efficacy. *Journal of Language Teaching and Research, 16*(6), 1848–1859. https://doi.org/10.17507/jltr.1606.06

Yeldham, M. (2018). Viewing L2 captioned videos: what's in it for the listener? *Computer Assisted Language Learning, 31*(4), 367–389. https://doi.org/10.1080/09588221.2017.1406956

Zadorozhnyy, A., & Lee, J. S. (2025). Informal digital learning of English and willingness to communicate in a second language: Self-efficacy beliefs as a mediator. *Computer Assisted Language Learning, 38*(4), 669–689. https://doi.org/10.1080/09588221.2023.2215279

Zaniar, S., Afandi, M. D., Aquariza, N. R., & Authar, N. (2025). Enhancing writing proficiency through mobile-assisted language learning (MALL): Insights and innovations. *English Language Teaching Methodology, 5*(1), 31–42. https://doi.org/10.56983/eltm.v5i1.1757

Zareian, G., Adel, S. M. R., & Noghani, F. A. (2015). The effect of multimodal presentation on EFL learners' listening comprehension and self-efficacy. *Academic Research International, 6*(1), Article 263. https://api.semanticscholar.org/CorpusID:219633047

Zhang, Q., Liu, X., & Chen, Y. (2021, November). Enhancing EFL learners' English speaking performance through vlog-based digital multimodal composing activities. In *International Symposium on Emerging Technologies for Education* (pp. 92–103). Springer International Publishing. https://doi.org/10.1007/978-3-030-92836-0_9

Zhao, Y., Li, Y., Ma, S., Xu, Z., & Zhang, B. (2025). A meta-analysis of the correlation between self-regulated learning strategies and academic performance in online and blended learning environments. *Computers & Education, 230*, Article 105279. https://doi.org/10.1016/j.compedu.2025.105279

Zheng, R., McAlack, M., Wilmes, B., Kohler-Evans, P., & Williamson, J. (2009). Effects of multimedia on cognitive load, self-efficacy, and multiple rule-based problem solving. *British Journal of Educational Technology, 40*(5), 790–803. https://doi.org/10.1111/j.1467-8535.2008.00859.x

Appendix A. Tasks of the Three Groups

Group type	Captioned vlog	Annotated vlog	Control group
Approach for Material Providing	Vlogs provided through Padlet. com		Audio files provided by the instructor
In-Class Tasks	50-minute mobile phone-based vlog learning task		50-minute traditional teacher-led task
	*__Stage One:__ 15-minute captioned vlog viewing adjusted to students' learning pace. *__Stage Two:__ 35-minute session to answer questions, provide feedback, and discuss.	* __Stage One:__ 15-minute annotated vlog viewing adjusted to students' learning pace. * __Stage Two:__ 35-minute session to answer questions, provide feedback, and discuss.	* __Stage One:__ 15-minute teacher-led lessons with audio files played by the instructor. * __Stage Two:__ 35-minute teacher-led session to answer questions and discuss responses with student group members.

Appendix B. Design Features of Captioned and Annotated Vlogs and Their Theoretical Rationale

Condition	Design Feature	Theoretical Principle	Expected Cognitive Effect
Captioned vlog	On-screen captions of spoken dialogue	Modality principle; redundancy management	Reinforcement of auditory input and support for spoken comprehension
Annotated vlog	Visual highlighting (color, boldface, framing) of target words	Signaling principle	Direction of attention to target vocabulary and facilitation of lexical noticing
	Bilingual gloss and illustrative image for target words	Multimedia principle; dual coding	Support for form-meaning mapping and semantic representation
	Annotation presented concurrently with the spoken word	Temporal contiguity principle	Integration of spoken input with explanatory information
	Annotation placed adjacent to the relevant dialogue	Spatial contiguity principle	Reduced split attention and more efficient processing
	Brief on-screen display of the annotation	Cognitive load management	Sufficient processing time while avoiding cognitive overload

Appendix C. Target Words in Short Conversations and Talks and Word Frequency

Word Frequency Results of Compleat VocabProfiler for the Short Conversations

Profile Summary		
K	#	cumulative(%)
K-1	3	10.7
K-2	10	46.4
K-3	9	78.6
K-4	3	89.3
K-5	2	96.4
K-6	1	100.0

allergy(K-5) option(K-2) process(K-2) payment(K-1) committee(K-2) venue(K-4) fail(K-2) convention(K-3) finalize(K-1) available(K-2) proposal(K-2) headquarters(K-3) strategy(K-3) campaign(K-3) launch(K-3) performance(K-2) conference(K-3) record(K-1) expire(K-5) receipt(K-4) previous(K-2) refund(K-6) current(K-2) estimate(K-3) lease(K-3) application(K-2) certificate(K-4) orientation(K-3)

Word Frequency Results of Compleat VocabProfiler for the Short Talks

Profile Summary		
K	#	cumulative(%)
K-1	3	10.7
K-2	8	39.3
K-3	12	82.1
K-4	1	85.7
K-5	2	92.9
K-6		-
K-7	1	96.4
K-8		-
K-9		-
K-10	1	100.0

engaging(K-2) invoice(K-7) division(K-3) accessory(K-5) complex(K-3) repair(K-2) install(K-3) competition(K-2) founder(K-3) depart(K-4) express(K-1) platform(K-3) disembark(K-10) apologize(K-3) delay(K-3) failure(K-3) annual(K-3) display(K-3) figure(K-1) predict(K-3) spot(K-1) intersection(K-5) surround(K-2) disturb(K-2) prohibit(K-3) wrap(K-2) purchase(K-2) assist(K-2)

Appendix D. Scales of Self-efficacy and Annotated Vlog Learning Perception Survey

Type	Item
Self-efficacy Scale	**Listening Self-efficacy** LSE 1. I believe I will receive a good grade by engaging in the listening tasks in this class. 我相信藉由參與本聽力任務可以獲得好的聽力成績。 LSE 2. I'm confident I can understand the content of the listening tasks in this course. 我有自信可以了解本聽力任務中的內容。 LSE 3. I'm confident I can answer listening questions regarding the listening tasks in this course.我有自信可以回答關於本聽力任務內容的問題。 LSE 4. I'm confident I can do an excellent job on the listening tasks in this course. 我有自信可以在本聽力任務中得到優良的表現。 LSE 5. I expect to do well in this listening tasks. 我預期我可以在本聽力任務中表現優良。 LSE 6. I'm certain I can master the listening skill via the listening tasks in this class. 我相信我可以透過本聽力任務來掌握聽力技巧。 **Vocabulary Self-efficacy** VSE 1. I'm certain I can understand the words shown in the listening tasks. 我相信我可以理解本聽力任務出現過的字詞。 VSE 2. I'm certain I can acquire new words shown in the listening tasks in this class. 我相信我可以透過本聽力任務來學習新單字。
Annotated Vlog Learning Perception Scale	1. Annotated vlogs can improve my English language. 附有註解影音部落格可以改善我的英文。 2. Annotated vlogs can enhance my English listening skills. 附有註解影音部落格可以提升我的英文聽力技巧。 3. Annotated vlogs can improve my English language pronunciation of words. 附有註解影音部落格可以改善我的英文單字發音。 4. I can learn new vocabulary from annotated vlogs. 我可以從附有註解影音部落格學習到新單字。 5. I can get used to annotated vlog speakers' speaking speed. 我可以習慣附有註解影音部落格中說話者的語速。 6. Watching annotated vlogs exposes me to English listening learning environment. 觀看附有註解影音部落格使我沉浸在英文聽力學習環境。 7. Annotated vlogs are far more interesting and engaging than the textbook videos and audio files. 附有註解影音部落格比課本影片及音檔，更有趣及更吸引人。 8. Annotated vlogs are good alternatives to the textbook videos and audio files. 與課本影片及音檔相比，附有註解影音部落格可以成為好的替代教材。 9. Annotated vlogs are suitable for use in the listening classes. 附有註解影音部落格適用於英文聽力課程。 10. Annotated vlogs can be used with different listening exercises and questions. 附有註解影音部落格可以搭配使用不同聽力練習和問題。

Appendix E. Four Open-ended Questions for the Semi-structured Interview

Item	Open-ended Question
1	Regarding the listening-learning task you participated in, what aspects did you feel were helpful to you? 針對您所參與的聽力學習任務，您覺得對您有幫助的部分是什麼?
2	Regarding the listening-learning task you participated in, what aspects did you feel were not helpful to you? 針對您所參與的聽力學習任務，您覺得對您沒有幫助的部分是什麼?
3	How is the listening-learning task you participated in different from the learning tasks you have experienced in the past? 您所參與的聽力學習任務與過去您所參與過的學習任務有何不同?
4	If there is an opportunity in the future, would you be willing to continue attending this listening-learning task? 若未來有機會的話，您還會有意願持續參與此聽力學習任務嗎?

About the Author

Dr. Hui-Tzu Hsu is an Assistant Professor at the Center for Language Studies, National Chung Cheng University in Taiwan. She obtained her Doctoral Degree of Philosophy in TESOL from National Taiwan Normal University. Her research interest primarily lies in computer-assisted language learning, mobile learning, vocabulary learning, learning motivation, second language acquisition, and English for specific purposes.

E-mail: lindahsu@ccu.edu.tw