Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Effects of Multimodal Glosses on Intentional Vocabulary Learning, Retention, and Learner Processing in Instagram Reels-Based Instruction

Raju Dhuli

raju.dhuli@shardauniversity.uz

Sharda University Uzbekistan

Saleem Mohd Nasim

Prince Sattam Bin Abdulaziz University

Sougand Akbarian

Islamic Azad University, Isfahan

Rakhimova Shahodat Azadovna

Tashkent Institute of Textile and Light Industry

Research Article

Keywords: Multimodal glosses, Instagram Reels, Intentional vocabulary learning, Vocabulary retention and Mobile-assisted language learning (MALL)

Posted Date: September 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11035031/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

The growing use of short-form social media platforms has created new opportunities for technology-enhanced vocabulary instruction in English as a Foreign Language (EFL) contexts. This study examined whether multimodal glosses embedded in Instagram Reels influenced intentional vocabulary learning, delayed retention, and self-reported cognitive engagement. A quasi-experimental pretest–posttest control-group design was conducted with 90 undergraduate EFL learners (45 experimental; 45 control). Both groups viewed the same 20 Instagram Reels containing identical audiovisual content and target vocabulary, while the experimental group received researcher-added multimodal glosses. Vocabulary knowledge was assessed using a Receptive and Contextual Vocabulary Knowledge Test (RCVKT) at pretest, immediate posttest, and two-week delayed posttest. An 18-item self-report questionnaire measured attention, lexical processing, and comprehension. Mixed repeated-measures ANOVA revealed a significant Group × Time interaction, with the experimental group outperforming the control group at both posttests, but not at pretest. Approximately 67% of immediate vocabulary gains were retained after two weeks. MANOVA showed significant group differences in self-reported attention and comprehension, but not lexical processing. The findings suggest that multimodal glosses in short-form video can support intentional vocabulary learning and retention, while also enhancing learners' reported attention and comprehension.

Introduction

Vocabulary knowledge is a fundamental component of second language (L2) proficiency, enabling learners to comprehend input, communicate effectively, and participate successfully in academic and social contexts. With the rapid growth of digital technologies, language learning has increasingly expanded beyond traditional classrooms to include social media platforms that provide authentic, engaging, and multimodal learning environments. Among these platforms, Instagram has gained considerable attention due to its ability to integrate text, images, audio, and video into short, visually appealing content that can facilitate contextualized vocabulary learning.

Instagram Reels, in particular, offer unique affordances for vocabulary instruction because they present lexical items within authentic communicative situations enriched by multiple semiotic resources. Their short-form format encourages repeated viewing, sustained learner attention, and exposure to vocabulary in meaningful contexts. Recent studies have shown that Instagram-based instruction, particularly through educational Reels, promotes learner engagement and positively influences vocabulary learning despite challenges such as rapid presentation and limited video duration (Aslan & Sirojitdinovna, 2025; La'Biran et al., 2024). Furthermore, Instagram's multimedia features support authentic language exposure while encouraging learner interaction through visual, textual, and social elements that contribute to vocabulary development and digital language learning experiences (Nasution, 2023; Sugiarto & Manara, 2025).

The effectiveness of Instagram Reels can be further enhanced through the integration of multimodal glosses. In this study, multimodal glosses refer to the combination of on-screen textual definitions with audiovisual contextual cues embedded within Instagram Reels. Their pedagogical value is grounded in multimodal learning principles, which propose that learners construct meaning more effectively when linguistic and visual information are processed together. By presenting vocabulary through complementary textual, visual, and auditory channels, multimodal glosses provide contextual support that facilitates lexical comprehension, strengthens form–meaning connections, and enhances delayed retention. Such multimodal coherence encourages learners to integrate multiple sources of information, resulting in richer cognitive processing and more durable vocabulary learning (Meier, 2022).

Intentional vocabulary learning refers to situations in which learners deliberately focus on acquiring new lexical items with the explicit goal of learning them. Research in second language acquisition has consistently demonstrated that multimedia resources, including captioned videos and glossed annotations, facilitate intentional vocabulary learning by simultaneously presenting lexical form, meaning, and contextual usage (Zhuo & Wang, 2025). Moreover, meta-analytic evidence indicates that multimedia glosses produce moderate to strong positive effects on vocabulary acquisition and retention, particularly among beginner and intermediate language learners by reducing cognitive demands and promoting deeper processing of new lexical items (PubMed Meta-analysis, 2024).

Although previous research has extensively investigated intentional vocabulary learning through captioned videos, digital stories, and other multimedia resources, relatively little attention has been devoted to Instagram Reels as an instructional medium. Unlike conventional educational videos, Instagram Reels combine brevity, authenticity, contextual richness, and multimodal presentation within a highly engaging social media environment. Emerging evidence suggests that these characteristics may enhance vocabulary learning by increasing learner motivation, attention, and repeated exposure to target words embedded in meaningful communicative contexts (Hasani et al., 2025). However, empirical research examining the specific contribution of on-screen meaning-based multimodal glosses within Instagram Reels remains limited. Consequently, it is still unclear whether such glosses facilitate not only immediate vocabulary acquisition but also delayed retention and learners' cognitive engagement.

Literature Review

Intentional Vocabulary Learning in Digital Environments

Vocabulary knowledge is fundamental to L2 development because it supports comprehension and production across language skills. However, vocabulary learning involves more than exposure; learners must notice, process, retrieve, and retain lexical information. In digital environments, extensive input does not necessarily ensure sufficient attention to target vocabulary. Thus, the value of technology depends less on its presence than on how its affordances direct attention to lexical form and meaning.

Intentional vocabulary learning (IVL) involves deliberate attention to selected lexical items. Digital environments can support IVL through repeated exposure, contextualization, and multimodal representations. Ramezanali et al. (2021) reported positive effects of technology-mediated vocabulary instruction, while Wang and Lee (2021) found advantages for multimedia-supported over text-only learning. However, these findings do not imply that technology itself produces vocabulary gains; rather, contextualization, visual representation, repetition, and accessibility may facilitate vocabulary learning.

Digital environments also present challenges. Muryani and Yunus (2024) reported benefits in accessibility and motivation alongside distraction and inconsistent learning habits, while Metruk and Zimenová (2025) highlighted flexibility, motivation, and learner autonomy in mobile-assisted vocabulary learning. Such findings indicate that engagement and accessibility do not necessarily translate into learning effectiveness. Teng (2024) further emphasizes learners' active attention to lexical information in multimedia environments. The key question, therefore, is not whether technology improves vocabulary learning generally, but which technological and instructional features make digital exposure cognitively productive.

Multimedia Learning and Vocabulary Acquisition

The Cognitive Theory of Multimedia Learning (CTML; Mayer, 2021) proposes that learners process verbal and visual information through partially distinct channels and may benefit when complementary representations are meaningfully integrated. In vocabulary learning, combining lexical form with visual, auditory, and contextual information may strengthen form–meaning connections.

Wang and Lee (2021) found stronger vocabulary outcomes for combined textual and visual representations than for text-only instruction. Such findings suggest that complementary representations, rather than the mere number of modes, facilitate learning when learners can meaningfully integrate them. Similarly, Teng (2024) reported positive effects of captioned videos on intentional vocabulary learning. These findings are also consistent with Dual Coding Theory (Paivio, 2006), which proposes that verbal and non-verbal representations can provide complementary routes for encoding and retrieval.

Nevertheless, multimodality does not automatically enhance learning. Simultaneous presentation of speech, captions, images, and animation may increase processing demands when information is redundant, irrelevant, or poorly synchronized. Mahdi et al. (2024), in a meta-analysis, found moderate-to-large positive effects of multimedia glosses on vocabulary acquisition and retention, but variation across studies indicates that effectiveness depends on learner characteristics, task design, and presentation conditions.

Thus, the literature distinguishes multimedia exposure from instructionally designed multimodal support. The present study focuses specifically on deliberately designed multimodal glosses embedded

within Instagram Reels to direct attention toward target vocabulary and support noticing, comprehension, processing, and retention.

Multimodal Glosses as an Instructional Intervention

Multimodal Glosses as an Instructional Intervention
Glosses traditionally provide immediate assistance with unfamiliar lexical items. Digital technologies have expanded glossing to include text, images, audio, animation, video, and contextual information, potentially creating richer form–meaning connections.

Ramezanali et al. (2021) found stronger vocabulary retention with multimodal than textual glosses, while Çekiç (2024) reported greater vocabulary gains under multimodal glossing during audiovisual learning. These findings suggest that multimodal glosses can support meaning construction and lexical processing. They can also function as attention-directing mechanisms by helping learners notice unfamiliar words and connect form, meaning, pronunciation, and contextual use, consistent with Schmidt's Noticing Hypothesis.

However, excessive or poorly designed information may increase cognitive demands. The effectiveness of glosses therefore depends on whether their components are complementary and relevant to the learning objective. Mahdi et al. (2024) provide broad evidence for the effectiveness of multimedia glosses, while Zhuo and Wang (2025) found that audiovisual glosses supported vocabulary acquisition by integrating pronunciation, meaning, and contextual usage.

Importantly, multimodal glosses should be distinguished from the platform through which they are delivered. In the present study, multimodal glosses constitute the pedagogical intervention, whereas Instagram Reels function as the technological delivery medium. This distinction helps identify whether learning gains are associated with the instructional design rather than with the platform itself.

Mobile-Assisted Language Learning as a Technological Environment

Mobile-Assisted Language Learning (MALL) provides flexible access to language resources through smartphones and applications. However, MALL is better understood as a technological environment or set of affordances than as a specific teaching method. It can support intentional and incidental learning, autonomous learning, collaborative activities, multimedia learning, and spaced practice.

Burston's (2015) meta-analysis established the potential of mobile technologies for language learning, while Metruk and Zimenová (2025) highlighted flexibility and motivational benefits in mobile-assisted vocabulary learning. Nevertheless, mobile technology itself does not constitute pedagogy. Its effectiveness depends on how its affordances are structured around learning objectives.

Mobile environments also provide opportunities for repeated exposure beyond classroom settings while creating potential distractions through notifications, entertainment, and competing information. Accordingly, Instagram Reels are conceptualized in this study as a mobile, short-form audiovisual delivery medium, rather than as a pedagogical method. The instructional intervention is the multimodal glossing embedded within the Reels.

Informal Digital Learning of English and Social Media

Informal Digital Learning of English (IDLE) involves digitally mediated English-learning activities outside conventional classroom instruction, including video viewing, social media participation, gaming, and online communication. Reynolds (2025) highlights the increasing role of informal digital practices in vocabulary development, while Baureh and Neman (2025) report learners' use of Instagram, TikTok, and YouTube for vocabulary development because of their authentic, contextualized, and multimodal input.

However, incidental exposure should be distinguished from intentional learning. Learners may encounter unfamiliar vocabulary through social media without deliberately attempting to learn it. Although repeated exposure creates learning opportunities, it does not necessarily result in durable lexical knowledge. Muryani and Yunus (2024) similarly identified both motivational benefits and distraction-related challenges in digital learning.

The present study therefore moves beyond social media exposure by incorporating an intentionally designed instructional treatment into a social media format. This approach combines the accessibility and engagement of informal digital environments with the explicit objectives of intentional vocabulary learning.

Instagram Reels as a Delivery Medium for Vocabulary Instruction

Instagram Reels combine spoken language, written text, images, demonstrations, and contextual scenes within short learning episodes, making them compatible with microlearning principles. Hasani et al. (2025) reported positive vocabulary-learning outcomes associated with Instagram Reels, while *Language Learning Made Short and Sweet? Exploring Student Perceptions of Microcelebrity Teacher Reels on Instagram* (2025) found that learners perceived teacher-created Reels as accessible and motivating.

However, emerging research has often emphasized engagement, accessibility, and learner perceptions rather than isolating the instructional mechanisms responsible for vocabulary learning. Improved performance after viewing Reels could result from repeated exposure, contextualized input, captions, visual information, novelty, motivation, or specific instructional support.

The present study therefore distinguishes two levels: Instagram Reels as the technological delivery medium and multimodal glosses as the pedagogical intervention. Rather than asking whether Reels are

generally effective, the study examines whether deliberately designed multimodal glosses delivered through Reels enhance intentional vocabulary learning and the cognitive processes associated with it.

Critical Synthesis and Research Gap

The literature supports three broad conclusions. First, digital and mobile environments can expand opportunities for vocabulary exposure, repetition, contextualization, and learner autonomy. Second, multimedia representations can facilitate vocabulary learning when they provide complementary support for form–meaning mapping. Third, appropriately designed glosses can promote learners' attention to and processing of unfamiliar lexical items.

However, technology, multimedia, and social media do not automatically produce vocabulary learning; outcomes depend on how technological affordances are combined with pedagogical design. In this context, MALL represents the technological environment, Instagram Reels the delivery medium, and multimodal glosses the instructional intervention. Distinguishing these levels is important for identifying the contribution of the instructional treatment.

A further gap concerns the limited integration of these strands of research. Multimedia learning, digital glossing, mobile applications, and social media have often been examined separately, and their findings cannot be assumed to generalize to short-form video environments characterized by brief exposure, rapid presentation, and high visual density. Moreover, social media research has frequently emphasized motivation, accessibility, and learner perceptions rather than measurable vocabulary acquisition and delayed retention.

Relatively little research has directly examined multimodal glosses embedded in Instagram Reels as an intentional vocabulary-learning intervention, particularly in relation to immediate learning, delayed retention, and learners' self-reported attention, lexical processing, and comprehension. Examining these outcomes together can provide a more comprehensive account of the potential contribution of multimodal glossing in short-form mobile learning.

Accordingly, the present study addresses a gap at the intersection of intentional vocabulary learning, multimodal glossing, mobile-assisted language learning, and short-form social media video. It investigates whether a specifically designed multimodal-gloss intervention delivered through Instagram Reels is associated with enhanced vocabulary acquisition and delayed retention, as well as differences in learners' self-reported attention, lexical processing, and comprehension.

Research Questions

1. To what extent does multimodal glossing enhance vocabulary learning from Instagram Reels compared with equivalent Reels without glossing?
2. How do learners' self-reported attention, lexical-processing engagement, and perceived comprehension differ between the multimodal-gloss and no-gloss conditions?

Methodology

Research Design

This study employed a quasi-experimental pretest–posttest control-group design. Two intact classes were assigned to the experimental and control conditions, respectively. Individual random assignment was not feasible because the participants remained in their pre-existing classes. The experimental group was exposed to Instagram Reels containing multimodal meaning-based visual glosses, whereas the control group viewed the same Reels without glosses.

Participants

The participants consisted of 90 BA second-year Applied English EFL university learners enrolled in undergraduate English courses at the Faculty of Humanities during the 2025–2026 academic year. The participants were aged 17–18 years. Uzbek was their primary language of communication outside the university, while English was the medium of instruction. The participants were drawn from two intact classes, with 45 learners in the experimental group and 45 learners in the control group. The two groups demonstrated comparable English proficiency, as indicated by their IELTS scores, which ranged from 6.0 to 6.5.

The experimental group received instruction through Instagram Reels containing multimodal meaning-based visual glosses, whereas the control group viewed the same Reels without glosses. Vocabulary performance was measured at three time points: pretest (Time 1), immediate posttest (Time 2), and delayed posttest (Time 3), with the delayed posttest administered two weeks after the immediate posttest. This design was used to examine vocabulary acquisition and longer-term retention (RQ1), as well as learners' attention, processing, and comprehension during exposure to the video materials (RQ2).

Materials

Instagram Reels as Instructional Materials

Instagram Reels served as the short-form, mobile-mediated delivery format rather than the intervention itself. Twenty researcher-selected Reels, each approximately 45–60 seconds long, were chosen based on four criteria: (a) linguistic suitability for the learners' proficiency level, (b) relevance and contextualizability of the target vocabulary, (c) availability of meaningful visual and situational information, and (d) technical clarity of the audiovisual input. The Reels contained authentic spoken English embedded in communicative contexts, with visual and situational information supporting interpretation of the target vocabulary.

The instructional objective was the intentional acquisition and subsequent retention of 100 unfamiliar vocabulary items. An initial pool of 250 words was identified from the selected Reels and screened using the Vocabulary Knowledge Scale (VKS; Wesche & Paribakht, 1996). Words rated as unfamiliar or minimally known (Levels 1–2) were considered for inclusion. Following screening for lexical frequency,

proficiency appropriateness, pedagogical learnability, and contextualizability, 100 target words were selected, with five target words assigned to each Reel.

The short, self-contained format provided focused episodes of contextualized audiovisual input while maintaining consistent exposure across materials. Importantly, both groups viewed the same 20 Reels with identical target vocabulary, audiovisual content, exposure duration, and contextual information. The experimental group additionally received researcher-added multimodal glosses comprising on-screen textual definitions, target-word highlighting, and contextual visual cues such as gestures, facial expressions, and situational information. The control group viewed the same Reels without these additional glosses. Thus, the presence or absence of the multimodal gloss package constituted the primary systematic difference between conditions. The study therefore examined the overall effect of multimodal glossing rather than the independent contribution of individual gloss components.

Vocabulary Knowledge Scale

The researcher employed the Vocabulary Knowledge Scale (VKS) to conduct an initial screening of participants' vocabulary knowledge in order to ensure that the used target words were unfamiliar to the learners prior to exposure to Instagram Reels. The VKS (Wesche & Paribakht, 1996) was administered to determine learners' prior knowledge of a pool of lexical items extracted from previously existing Reels. For this purpose, 250 words were selected and presented to the participants through the VKS instrument. Target words were selected from the mid-frequency range of the British National Corpus (BNC), screened for unfamiliarity using the VKS, and further evaluated for proficiency appropriateness, pedagogical learnability, and contextualizability.

Based on participants' responses, it was found that 100 out of the 250 words were unfamiliar to the learners, as indicated by responses corresponding to levels 1 and 2 of the VKS scale (i.e., "I have never encountered this word before" or "I have seen this word but do not know its meaning"). These 100 unfamiliar words were subsequently selected as target vocabulary items for the main intervention. This procedure ensured that the instructional materials focused on genuinely unknown vocabulary, thereby enabling a valid assessment of intentional vocabulary learning through the treatment.

Development, Validation, and Item Analysis of the Receptive and Contextual Vocabulary Knowledge Test (RCVKT)

The Receptive and Contextual Vocabulary Knowledge Test (RCVKT; see Appendix A) was developed to assess participants' knowledge of the 100 target vocabulary items encountered during the Instagram Reels intervention. It was administered at the pretest, immediate posttest, and delayed posttest stages using three task formats: multiple-choice meaning recognition, form–meaning matching, and sentence completion. Thus, the 100 target words generated 300 dichotomously scored responses. Correct responses received one point and incorrect or unanswered responses zero, yielding a maximum score of

300, with higher scores indicating greater vocabulary knowledge. Parallel forms were used across the three testing points to minimize guessing and test-practice effects.

For content validity, two experts in English language teaching and vocabulary assessment reviewed the initial RCVKT for item relevance and clarity, task-format appropriateness, distractor quality, and alignment with the intended vocabulary-learning objectives. Ambiguous or poorly aligned items were revised before pilot testing.

The revised RCVKT was pilot-tested with 40 EFL learners comparable in proficiency to the main sample. Item difficulty was assessed using the proportion-correct index (p), with higher values indicating easier items (Crocker & Algina, 1986; Rezigalla et al., 2024). Item discrimination was examined using the point-biserial correlation (rpb), with weak or negative coefficients reviewed for ambiguity, inappropriate distractors, extreme difficulty, scoring errors, or answer-keying problems (Ebel & Frisbie, 1991; Rezigalla et al., 2024). For the multiple-choice items, distractor functioning was also examined; distractors selected by fewer than 5% of participants were considered potentially non-functioning and reviewed for revision (Haladyna et al., 2002; Rezigalla et al., 2024).

Following item analysis and revision, internal consistency was assessed using the Kuder–Richardson Formula 20 (KR-20), given the dichotomous scoring of the 300 responses. The final RCVKT demonstrated good internal consistency (KR-20 = .82). Reliability was also examined separately for the three task formats. The same target vocabulary, task formats, scoring procedures, and administration conditions were maintained across all three testing points to ensure consistency in measuring vocabulary development and delayed retention.

Operationalization of the Multimodal Gloss

In the present study, multimodal glossing was operationalized as an integrated package of researcher-added meaning-support features embedded within the Instagram Reels. Both groups received the same spoken target words, natural video context, and on-screen presentation of the target vocabulary. In the experimental condition, the target words were additionally supported through brief textual definitions, and keyword highlighting. These features were designed to provide immediate meaning support and facilitate learners' form–meaning connections during video viewing.

The control group viewed the same underlying Reels without the researcher-added glossing features. Thus, the two conditions were equivalent in terms of video content, target vocabulary, audiovisual context, exposure time, and instructional setting, differing only in the presence or absence of the multimodal gloss package. The experimental manipulation therefore concerned the combined effect of the multimodal gloss package as a whole. Because all gloss components were presented simultaneously, the study was not designed to determine the independent contribution of textual definitions, keyword highlighting, supplementary images, or animation. Accordingly, the findings are interpreted in terms of the effectiveness of the multimodal gloss package rather than the causal effect of any individual component.

Attention, Processing, and Comprehension Measures (RQ2)

a. Learner Perception Questionnaire

An 18-item self-report questionnaire was developed to examine learners' perceived engagement with Instagram Reels and their reported responses to the vocabulary-learning activities. It comprised three six-item dimensions: (a) self-reported attentional engagement, (b) self-reported lexical processing, and (c) self-reported video comprehension. Items were rated on a five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree), with higher scores indicating greater levels of the respective self-reported construct.

Self-reported attentional engagement assessed learners' reported attention to video content and target vocabulary. Self-reported lexical processing assessed reported learning behaviors such as inferring meanings from context, mentally repeating new words, connecting vocabulary with prior knowledge, and attempting to remember word forms and meanings. Self-reported video comprehension assessed learners' perceived understanding of the Reels' content and meaning.

Two experts reviewed the questionnaire for item relevance, clarity, wording, and alignment with the intended dimensions. Internal consistency was assessed using Cronbach's alpha. The overall questionnaire showed good reliability ($\alpha = .866$), with coefficients of .835 for attentional engagement, .741 for lexical processing, and .756 for video comprehension.

Because the questionnaire relied on self-report, its scores were interpreted as learners' perceptions and reported engagement rather than direct measures of attention or cognitive processing. In particular, the lexical-processing dimension reflects perceived processing activities rather than objectively measured processing depth or quality, while attentional engagement reflects perceived rather than observed attention. Thus, the questionnaire provides evidence of learners' reported engagement and perceptions, not direct evidence of attention, gaze behavior, or underlying cognitive processing.

Pilot Testing

Comprehension Test

A short video comprehension test was used to assess understanding of the reel content, ensuring that vocabulary learning did not occur at the expense of overall comprehension.

Data Collection Procedure

1. Orientation Session

1. Participants were informed that the study focused on general video comprehension; vocabulary learning was mentioned to ensure intentional learning conditions.

3. Pretesting (Time 1)

1. All participants completed a vocabulary pretest to confirm minimal prior knowledge of the target words.

Intervention and Assessment Procedure

The intervention was conducted over three weeks and comprised 10 instructional sessions of 60 minutes each, with two Instagram Reels presented per session. Thus, participants were exposed to 20 Reels containing 100 target vocabulary items. The experimental group viewed the Reels with researcher-added multimodal glosses designed to support attention to, processing of, and learning of the target vocabulary. The control group viewed the identical Reels without glosses and was instructed to watch them naturally without note-taking or dictionary use. Both groups received the same instructional time, Reels, target vocabulary, audiovisual content, and classroom conditions; the presence or absence of multimodal glosses constituted the primary systematic difference between the groups.

Immediate Posttest (Time 2)

Immediately after the 10-session intervention, participants completed the vocabulary posttest and learner perception questionnaire.

Delayed Posttest (Time 3)

Two weeks after the immediate posttest, participants completed the delayed vocabulary posttest to assess retention of the target vocabulary.

Data Analysis

All analyses were conducted using IBM SPSS Statistics Version 28 at $\alpha = .05$ (two-tailed). Data were screened for missing values, outliers, normality, homogeneity of variance, and sphericity, with appropriate corrections applied when assumptions were violated.

For RQ1, a 2 × 3 mixed repeated-measures ANOVA examined RCVKT performance across pretest, immediate posttest, and delayed posttest between the experimental and control groups. Significant Group × Time effects were followed by Bonferroni-adjusted pairwise comparisons. Pillai's Trace was used as a robustness check, and partial eta squared (ηp^2) was reported. Within-group pre–post gains were examined using paired-samples *t*-tests, while delayed vocabulary maintenance was assessed using a retention index and between-group comparison. Cohen's *d* and 95% confidence intervals were reported where appropriate.

For RQ2, a one-way MANOVA examined group differences in self-reported attention, lexical processing, and comprehension, followed by univariate analyses where warranted. Effect sizes were reported using partial eta squared (ηp^2).

Ethical considerations

This study involved mobile based standard classroom reading activities and assessments. Participation was voluntary, and all participants provided informed consent prior to the study. Data were treated confidentially and reported in aggregate form. As the research involved routine educational activities with no foreseeable risk to participants, formal ethics committee approval was not required.

Results

RQ1: Impact of Multimodal Glosses on Intentional Vocabulary Acquisition and Retention

Table 4
Descriptive Statistics for Vocabulary Performance (RCVKT Scores)

Group	Pretest M (SD)	Posttest M (SD)	Delayed Posttest M (SD)
Control	78.33 (7.32)	78.60 (6.45)	77.89 (7.42)
Experimental	80.09 (5.37)	135.27 (17.08)	117.27 (17.08)
Total	79.21 (6.44)	106.93 (31.25)	97.58 (23.74)

Table 4 presents the descriptive statistics of RCVKT scores across pretest, posttest, and delayed posttest conditions for both experimental and control groups. At pretest, both groups demonstrated comparable vocabulary knowledge, with minimal difference between the control group (M = 78.33, SD = 7.32) and the experimental group (M = 80.09, SD = 5.37), indicating baseline equivalence prior to treatment.

Following the intervention, a sharp divergence was observed. The experimental group showed a substantial increase in vocabulary performance at posttest (M = 135.27, SD = 17.08), whereas the control group remained nearly unchanged (M = 78.60, SD = 6.45). At the delayed posttest, the experimental group experienced a decline (M = 117.27, SD = 17.08), but their performance remained substantially higher than both their pretest level and the control group's delayed performance (M = 77.89, SD = 7.42). These patterns indicate strong immediate learning gains and partial retention of vocabulary knowledge in the experimental condition.

Table 5
Mixed ANOVA Results for Time × Group Interaction

Effect	F	df	p	Partial η^2
Time	240.88	(1.27, 111.87)	< .001	.732
Time × Group	238.66	(1.27, 111.87)	< .001	.731

Table 5 reports the within-subjects effects of time and the interaction effect between time and group. The results reveal a statistically significant main effect of time, $F(1.27, 111.87) = 240.88$, $p < .001$, with a large effect size ($\eta p^2 = .732$), indicating that participants' vocabulary scores changed significantly across the three testing occasions.

More importantly, a statistically significant Time × Group interaction was observed, F(1.27, 111.87) = 238.66, p < .001, ηp² = .731. This indicates that the pattern of vocabulary development over time differed significantly between the experimental and control groups. The magnitude of the effect suggests that the interaction showed a very large effect (ηp² = .731), indicating substantial differences in vocabulary change over time between the two groups., reflecting a strong instructional effect.

Table 6
Between-Subjects Effect of Group

Effect	F	df	p	Partial η²
Group	308.29	(1, 88)	< .001	.778

Table 6 presents the between-subjects effect of group membership on overall vocabulary performance. The results indicate a statistically significant main effect of group, F(1, 88) = 308.29, p < .001, ηp² = .778. This finding confirms that there was a substantial difference in vocabulary performance between learners exposed to multimodal glosses and those in the control condition.

The effect size indicates that approximately 77.8% of the variance in overall vocabulary scores is explained by group membership, demonstrating a very strong advantage for the experimental group. This result provides strong evidence that multimodal glossing in Instagram Reels had a powerful impact on learners' intentional vocabulary acquisition.

Table 7
Bonferroni-adjusted pairwise comparisons of Time within each group

Group	Comparison (I − J)	Mean Difference	SE	Adjusted *p*
Control	Pretest − Immediate Posttest	−0.27	2.12	1.000
	Pretest − Delayed Posttest	0.44	2.15	1.000
	Immediate Posttest − Delayed Posttest	0.71	0.90	1.000
Experimental	Pretest − Immediate Posttest	−55.18	2.12	< .001
	Pretest − Delayed Posttest	−37.18	2.15	< .001
	Immediate Posttest − Delayed Posttest	18.00	0.90	< .001

Note. Mean differences are calculated as I − J. *p* values are Bonferroni-adjusted. The comparisons use the error term from the repeated-measures GLM.

Bonferroni-adjusted pairwise comparisons showed no significant differences across any time points in the control group (*p* = 1.000). In contrast, all three comparisons were significant in the experimental group: pretest–immediate posttest (MD = − 55.18, *p* < .001), pretest–delayed posttest (MD = − 37.18, *p* < .001), and immediate–delayed posttest (MD = 18.00, *p* < .001). Although performance declined during

the two-week delay, the delayed-posttest mean (117.27) remained substantially above the pretest mean (80.09), indicating partial retention of vocabulary gains.

Table 8
Bonferroni-adjusted pairwise comparisons of Groups at each time point

Measurement occasion	Comparison (Control – Experimental)	Mean Difference	SE	Adjusted *p*
Pretest	Control – Experimental	−1.76	1.35	.594
Immediate Posttest	Control – Experimental	−56.67	2.72	< .001
Delayed Posttest	Control – Experimental	−39.38	2.78	< .001

The Bonferroni-adjusted comparisons showed that the control and experimental groups did not differ significantly at pretest (MD = − 1.76, adjusted p = .594), confirming comparable baseline vocabulary performance. However, the experimental group significantly outperformed the control group at the immediate posttest (MD = − 56.67, adjusted p < .001). The between-group difference remained significant at the delayed posttest (MD = − 39.38, adjusted p < .001). These findings indicate that the experimental group's advantage emerged following the intervention and was maintained two weeks later.

Table 9
Sphericity Test (Mauchly's Test)

Effect	W	χ^2	df	p
Time	.427	74.088	2	< .001

Table 9 presents Mauchly's Test of Sphericity for the within-subjects factor (time). The results indicate that the assumption of sphericity was violated, W = .427, $\chi^2(2) = 74.088$, p < .001. Therefore, Greenhouse–Geisser corrections were applied in the interpretation of the repeated-measures ANOVA results. Despite this violation, the corrected results remained statistically significant, confirming the robustness of the findings.

Table 10
Retention Index for Vocabulary Maintenance

Group	Pretest	Posttest	Delayed Posttest	Retention Index (%)
Experimental	80.09	135.27	117.27	**67.38**
Control	78.33	78.60	77.89	Not interpreted*

Table 10 presents the Retention Index for vocabulary maintenance. The experimental group achieved a Retention Index of 67.38%, indicating that approximately two-thirds of the vocabulary gains observed immediately after the intervention were retained two weeks later. This finding demonstrates that although some decline occurred between the immediate and delayed posttests, a substantial proportion

of the acquired vocabulary knowledge was maintained over time. The Retention Index was not interpreted for the control group because the negligible improvement between the pretest and posttest resulted in an unstable denominator, making the index unsuitable for meaningful comparison.

Research Question 2: To what extent does multimodal glossing affect learners' self-reported attentional engagement, self-reported lexical processing, and self-reported video comprehension?

Table 11
Descriptive Statistics for Attention, Processing, and Comprehension by Group

Variable	Group	Mean	SD	N
Attentional engagement	Control	3.47	0.77	45
	Experimental	4.41	0.48	45
lexical processing	Control	3.33	0.72	45
	Experimental	3.43	0.67	45
Video Comprehension	Control	3.36	0.65	45
	Experimental	4.27	0.51	45

Table 11 presents descriptive statistics for learners' self-reported attentional engagement, lexical processing, and video comprehension. The experimental group reported higher attentional engagement (M = 4.41, SD = 0.48) and video comprehension (M = 4.27, SD = 0.51) than the control group (M = 3.47, SD = 0.77; M = 3.36, SD = 0.65, respectively). In contrast, self-reported lexical processing differed only marginally between the experimental (M = 3.43, SD = 0.67) and control (M = 3.33, SD = 0.72) groups. Overall, the experimental group reported greater attentional engagement and perceived comprehension, while the groups showed similar levels of self-reported lexical processing.

Table 12
Multivariate Test of the Effect of Group on Attention, Processing, and Comprehension

Effect	Test	Value	F	df (Hypothesis, Error)	p	Partial η^2
Group	Pillai's Trace	.419	20.639	(3, 86)	< .001	.419

Table 12 presents the multivariate analysis of variance (MANOVA) examining the overall effect of multimodal glosses on learners' self-reported attentional engagement, lexical processing, and comprehension. The results revealed a statistically significant multivariate effect of group, Pillai's Trace = .419, $F(3, 86) = 20.639$, $p < .001$, partial $\eta^2 = .419$. This finding indicates that the combined dependent variables differed significantly between the experimental and control groups. The effect size indicated a large multivariate group effect, suggesting a substantial difference between the two groups across the combined measures.

Table 13
Levene's Test of Equality of Error Variances

Variable	F	df1	df2	p
Attentional engagement	7.183	1	88	.009
lexical processing	0.364	1	88	.548
Video Comprehension	1.909	1	88	.171

Levene's test indicated that the homogeneity-of-variance assumption was satisfied for self-reported lexical processing (p = .548) and comprehension (p = .171), but violated for attention (p = .009). Because the groups were equal in size (n = 45 each), the MANOVA is reasonably robust to moderate variance heterogeneity. Nevertheless, the multivariate findings should be interpreted cautiously, particularly for attention.

Table 14
Tests of Between-Subjects Effects

Dependent Variable	F	p	Partial η²
Attentional engagement	48.143	< .001	.354
lexical processing	0.507	.478	.006
Video Comprehension	53.592	< .001	.378

Table 14 presents the univariate effects of group on each dependent variable. Significant group effects were found for self-reported attentional engagement, $F(1, 88) = 48.143$, $p < .001$, partial $\eta^2 = .354$, and self-reported comprehension, $F(1, 88) = 53.592$, $p < .001$, partial $\eta^2 = .378$. Both effects were large, with the multimodal-gloss group reporting higher attention and comprehension than the control group. In contrast, self-reported lexical processing did not differ significantly between groups, $F(1, 88) = 0.507$, p = .478, partial $\eta^2 = .006$, indicating a negligible group effect.

Discussion

The present study examined the effectiveness of multimodal glosses embedded in Instagram Reels on learners' intentional vocabulary learning, delayed retention, and cognitive engagement during video-based language exposure. The findings indicate that learners exposed to multimodal glosses demonstrated substantially greater vocabulary development than learners who viewed the same reels without glosses. The significant Group × Time interaction, together with the follow-up simple-effects and pairwise comparisons, provides evidence that the two groups developed differently across the three measurement occasions. In particular, the experimental group showed substantial improvement from pretest to immediate posttest and maintained significantly higher vocabulary performance at the two-week delayed posttest, whereas the control group remained statistically stable.

Effect of Multimodal Glosses on Vocabulary Learning and Delayed Retention

The results for RQ1 revealed clear differences between the two groups over time. The groups did not differ significantly at pretest ($p = .198$), indicating comparable initial vocabulary performance. Following the intervention, the experimental group significantly outperformed the control group at both the immediate ($p < .001$) and two-week delayed posttests ($p < .001$). The significant Group × Time interaction reflected substantial improvement in the experimental group, whereas the control group remained relatively stable.

The experimental group increased from a mean of 80.09 at pretest to 135.27 at immediate posttest and 117.27 at delayed posttest. Although performance declined significantly over the two-week interval, delayed scores remained substantially above pretest levels, with 67.38% of immediate gains retained. In contrast, the control group showed no significant change, with means of 78.33, 78.60, and 77.89, respectively. This contrasting pattern suggests that multimodal gloss exposure was associated with greater vocabulary learning and short-term retention than exposure to the same Reels without gloss support.

These findings are consistent with research showing that multimedia-enhanced vocabulary instruction can facilitate form–meaning learning through multiple channels of meaning representation (Ramezanali et al., 2020; Wang & Lee, 2021). They also accord with the meta-analytic findings of Mahdi et al. (2024), which reported moderate to strong effects of visual and multimedia glosses on vocabulary learning in instructed contexts.

The findings can be interpreted through Dual Coding Theory (Paivio, 2006) and the Cognitive Theory of Multimedia Learning (Mayer, 2021). The combination of verbal input, visual highlighting, and contextualized video information may have supported interconnected representations and form–meaning associations. From a multimedia-learning perspective, learners may also have been better able to select and integrate complementary verbal and visual information. However, because the study did not directly measure the cognitive processes underlying learning, these explanations should be regarded as plausible interpretations rather than demonstrated mechanisms.

The significant decline from immediate to delayed posttest indicates that the initial gains were not fully maintained over two weeks. Thus, multimodal glosses may support initial vocabulary learning, but additional reinforcement may be needed to strengthen longer-term retention. This interpretation is consistent with Teng (2024), who emphasizes repeated exposure and retrieval for maintaining vocabulary knowledge in digital learning environments. The 67.38% retention index further indicates that a substantial proportion of the immediate learning gains was maintained after two weeks.

Influence of Multimodal Glosses on Self-Reported Attentional Engagement, Lexical Processing, and Comprehension

The results for RQ2 showed significant group differences in self-reported attentional engagement and comprehension, but not in self-reported lexical processing. The largest effect was observed for comprehension (partial $\eta^2 = .378$), followed by attentional engagement (partial $\eta^2 = .354$), whereas lexical processing showed a negligible effect (partial $\eta^2 = .006$).

The strong effect for self-reported comprehension suggests that multimodal glosses may have provided useful meaning support during video exposure by clarifying unfamiliar vocabulary and facilitating learners' perceived understanding of the content. This interpretation is broadly consistent with Mayer's (2021) Cognitive Theory of Multimedia Learning, which proposes that appropriately integrated verbal and visual information can support meaningful learning. However, because comprehension was measured through self-report, the finding reflects perceived comprehension rather than objectively measured comprehension performance.

The significant effect on self-reported attentional engagement suggests that the visual salience of the glossed vocabulary may have increased learners' reported focus on target lexical items. This interpretation is compatible with Schmidt's Noticing Hypothesis, which emphasizes attention to linguistic input as an important condition for learning. It is also broadly consistent with research reporting benefits of visual annotations for attention and vocabulary learning in multimedia environments (Ramezanali et al., 2021; Çekiç, 2024). Nevertheless, the present study did not directly measure attention or noticing; therefore, these mechanisms remain plausible interpretations rather than demonstrated processes.

In contrast, self-reported lexical processing did not differ significantly between groups. This finding suggests that multimodal glosses, while associated with higher reported attention and comprehension, were not sufficient to produce differences in learners' reported lexical-processing activities. More elaborative processing may require additional opportunities for semantic analysis, retrieval, repetition, and active manipulation of new vocabulary. From a Levels of Processing perspective (Craik & Lockhart, 1972), such elaborative engagement may contribute to stronger and more durable memory representations.

Importantly, the non-significant lexical-processing effect should not be interpreted as evidence that multimodal glosses had no influence on actual cognitive processing. Rather, it indicates that no group difference was detected on the self-report measure used to assess lexical processing. Direct claims about cognitive processing would require behavioral or process-tracing measures, such as response-time measures, eye-tracking, or think-aloud protocols.

Relationship between Cognitive Engagement and Vocabulary Learning

The findings concerning cognitive engagement should be interpreted cautiously. Although the experimental group demonstrated substantial vocabulary gains and significant effects were observed for attention and comprehension, the present study did not directly test a causal relationship between these cognitive dimensions and vocabulary development. Therefore, it would be inappropriate to conclude that attention or comprehension independently caused the vocabulary gains.

Nevertheless, the pattern of findings suggests that enhanced noticing and comprehension may have contributed to successful vocabulary learning under the multimodal-gloss condition. Learners were exposed to target vocabulary in meaningful video contexts while receiving immediate visual meaning support. This combination may have facilitated initial recognition and comprehension of target lexical items even when opportunities for deeper elaborative processing were limited.

This finding is broadly consistent with research in mobile-assisted language learning, which suggests that vocabulary learning in digital environments can be supported through meaningful exposure, contextualization, and multimodal input rather than through explicit memorization alone (Kukulska-Hulme & Shield, 2008; Reinders & Benson, 2017). However, the present findings also indicate that such exposure may benefit from complementary activities designed to encourage deeper processing and retrieval.

Pedagogical Implications

The findings suggest that multimodal glosses embedded in Instagram Reels can support intentional vocabulary learning, particularly learners' self-reported attention and comprehension. Teachers may therefore incorporate multimodal glosses into short-form videos to provide contextualized meaning support. However, the decline from immediate to delayed posttest indicates that initial gains were not fully maintained. Glossed video exposure should therefore be supplemented with retrieval practice, spaced repetition, vocabulary recycling, sentence production, and meaningful use to strengthen longer-term retention.

The limited change in self-reported lexical processing further suggests that multimodal glosses alone may not promote deeper lexical engagement. Teachers could complement glossed videos with activities requiring learners to analyze meanings, compare lexical forms, retrieve target words, generate sentences, and use new vocabulary in meaningful contexts.

Limitations and Conclusion

This study provides evidence that multimodal glosses embedded in Instagram Reels can support intentional vocabulary learning, but several limitations should be acknowledged. The sample was limited to 90 undergraduate EFL learners from a single university in Uzbekistan, restricting generalizability across educational, linguistic, and cultural contexts. The three-week intervention and two-week delayed posttest also provide limited evidence of long-term retention; future studies should employ longer interventions and extended follow-up assessments. In addition, the RCVKT measured receptive and

contextual vocabulary knowledge rather than productive use. Future research should therefore examine whether Reels-based vocabulary learning transfers to speaking and writing.

Although the study examined learners' self-reported attention, lexical processing, and comprehension, it did not include objective measures such as eye-tracking, screen-recording analytics, or think-aloud protocols. Thus, the findings should not be interpreted as direct evidence of enhanced cognitive processing. Future research could combine self-report with objective measures and compare different gloss types, including textual, pictorial, audio, animated, and interactive formats. Learner proficiency, cognitive load, motivation, and other learner-related factors could also be examined as potential moderators.

Overall, learners exposed to multimodal glosses achieved significantly greater vocabulary gains than those viewing the same Reels without glosses, with approximately 67% of immediate gains retained after two weeks. Multimodal glosses also produced higher self-reported attention and comprehension, but not lexical processing. These findings are broadly consistent with the Cognitive Theory of Multimedia Learning, Dual Coding Theory, and the Noticing Hypothesis, although the non-significant lexical-processing effect prevents claims that deeper processing explains the vocabulary gains. The study extends MALL and IDLE research by demonstrating the potential of Instagram Reels for contextualized vocabulary instruction. Pedagogically, multimodal Reels may be strengthened through retrieval practice, vocabulary recycling, spaced repetition, and contextualized production to promote longer-term retention and lexical use.

Declaration of AI Use

The authors used generative artificial intelligence tools (ChatGPT, developed by OpenAI) to assist with language editing and stylistic refinement of this manuscript. All AI-assisted outputs were reviewed, verified, and revised by the authors. The authors retain full responsibility for the accuracy, originality, and integrity of the content, as well as for the interpretations and conclusions presented in this work.

Declarations

Author Contribution

Author Contribution DeclarationRaju Dhuli contributed to the conceptualization, methodology, investigation, development of study materials, data collection, data analysis, and writing of the manuscript, including the original draft, review, and editing. Saleem Mohd Nasim contributed to the conceptualization, methodology, supervision, critical review, and editing of the manuscript. Sougand Akbarian and Rakhimova Shahodat Azadovna contributed to the investigation, data collection, development of study materials, critical review, and editing of the manuscript. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work.

References

1. Aslan, E., & Sirojitdinovna, M. B. (2025). Language learning made short and sweet? Exploring student perceptions of microcelebrity teacher reels on Instagram. *Linguistics and Education*, *88*, 101430. https://doi.org/10.1016/j.linged.2025.101430
2. Baureh, G. A. D., & Neman, M. I. E. (2025). Learners' Perceptions of Social Media-Based Vocabulary Learning: Insights from TikTok, Instagram, and YouTube Users in an EFL Context. *Journal of English Language and Education*, *10*(6), 668–673. https://doi.org/10.31004/jele.v10i6.1742
3. Burston, J. (2015). Twenty years of MALL project implementation: A meta-analysis. *ReCALL*, *27*(1), 4–20. https://doi.org/10.1017/S0958344014000159
4. Çekiç, A. (2024). Intentional L2 vocabulary learning from audiovisual input: The effects of different types of glosses. *Computer Assisted Language Learning*, *37*(4), 896–923. https://doi.org/10.1080/09588221.2022.2037724
5. Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt. 6277 Sea Harbor Drive, Orlando, FL 32887.
6. Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Prentice Hall.
7. Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied measurement in education*, *15*(3), 309–333. https://doi.org/10.1207/S15324818AME1503_5
8. Hasani, M. M., Zainuddin, Z., & Hamdani, B. (2025). The influence of Instagram Reels on English vocabulary acquisition among English learners. *International Journal of English Education and Linguistics (IJoEEL)*, *7*(1), 139–146. https://doi.org/10.33650/ijoeel.v7i1.11295
9. Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning. *ReCALL*, *20*(3), 271–289. https://doi.org/10.1017/S0958344008000335
10. La'Biran, R., Ruung, T., T., & Sallata, Y. N. (2024). Exploring the Influence of Instagram Usage on English Education Students' Vocabulary Acquisition at UKI Toraja. *International Journal of Social Science and Human Research*, *07*(01). https://doi.org/10.47191/ijsshr/v7-i01-62
11. Mahdi, H. S., Mohsen, M. A., & Almanea, M. (2024). Multimedia glosses and second language vocabulary learning: A meta-analysis. *Acta Psychologica*, *248*, 104341. https://doi.org/10.1016/j.actpsy.2024.104341
12. Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press.
13. Meier, S. (2022). Digital Storytelling: A Didactic Approach to Multimodal Coherence. *Frontiers in Communication*, *7*. https://doi.org/10.3389/fcomm.2022.906268
14. Metruk, R., & Zimenová, I. (2025). The use of digital technologies in the process of learning vocabulary. *Information Technologies and Learning Tools*, *109*(5), 69–84. https://doi.org/10.33407/itlt.v109i5.6284
15. Muryani, R. N., & Yunus, M. (2024). Students 'digital vocabulary learning: experiences, barriers, and adaptive strategies exploration. *English Review: Journal of English Education*, *12*(1), 31–42.

https://doi.org/10.25134/erjee.v12i1.9328

16. Nasution, A. K. P. (2023). Instagram in English Language Learning: A Systematic Literature Review. *Journal of Linguistics Literature and Language Teaching (JLLLT)*, *3*(1), 33–52. https://doi.org/10.37249/jlllt.v3i1.708
17. Paivio, A. (2006). *Mind and its evolution: A dual coding theoretical approach*. Psychology.
18. Ramezanali, N., Uchihara, T., & Faez, F. (2021). Efficacy of multimodal glossing on second language vocabulary learning: A meta-analysis. *Tesol Quarterly*, *55*(1), 105–133. https://doi.org/10.1002/tesq.579
19. Reinders, H., & Benson, P. (2017). Research agenda: Language learning beyond the classroom. *Language Teaching*, *50*(4), 561–578. https://doi.org/10.1017/S0261444817000192
20. Ren, C., & Su, L. (2025). The Role of Digital Tools in Enhancing Vocabulary Acquisition in Second Foreign Language Learning. *Journal of Contemporary Educational Research*, *9*(7), 376–381.
21. Reynolds, B. L. (2025). Informal digital learning of English and vocabulary development. *Education Sciences*, *15*(12), 1606. https://doi.org/10.3390/educsci15121606
22. Rezigalla, A. A., Eleragi, A. M., et al. (2024). Item analysis: the impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *Bmc Medical Education*, *24*, 445. https://doi.org/10.1186/s12909-024-05433-y
23. Sugiarto, B. R., & Manara, C. (2026). Digital multilingual identity: insights from Indonesian multilingual EFL students' identity construction on Instagram. *International journal of multilingualism*, *23*(1), 1021–1042. https://doi.org/10.1080/14790718.2025.2508867
24. Teng, M. F. (2024). Intentional vocabulary learning from reading, listening, and viewing captioned videos. *System*, *124*, 103381. https://doi.org/10.1016/j.system.2024.103381
25. Teng, M. F. (2024). Working memory and prior vocabulary knowledge in incidental vocabulary learning from listening, reading, reading-while-listening, and viewing captioned videos. *System*, *124*, 103381. https://doi.org/10.1016/j.system.2024.103381
26. Wang, Y., & Lee, J. (2021). Multimedia gloss presentation and EFL vocabulary learning. *Frontiers in Psychology*, *11*, 602520. https://doi.org/10.3389/fpsyg.2020.602520
27. Wang, S., & Lee, C. I. (2021). Multimedia gloss presentation: Learners' preference and the effects on EFL vocabulary learning and reading comprehension. *Frontiers in Psychology*, *11*, 602520. https://doi.org/10.3389/fpsyg.2020.602520
28. Wang, Z., Zou, D., Peng, P., Wang, F. L., Lee, L. K., & Xie, H. (2025). Effects of mobile metaverse-based vocabulary learning on learners' perception and performance: a case study of Chinese EFL learners. *Journal of Computers in Education*, *12*(4), 1333–1366. https://doi.org/10.1007/s40692-024-00348-5

Figures

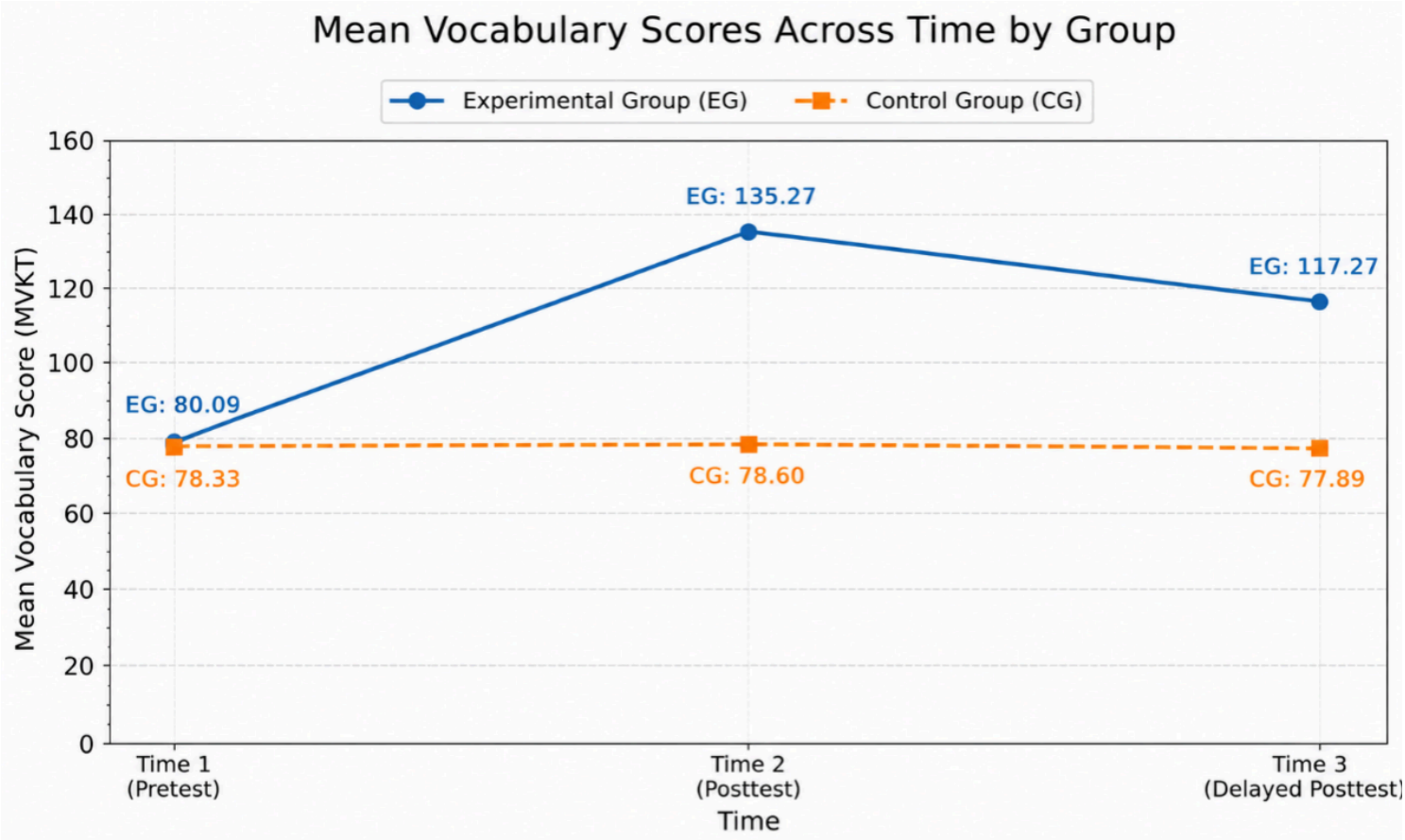


Figure 1

Mean Vocabulary Scores Across Time by Group (Experimental vs. Control)

Note. Scores represent total correct responses on the 0–300 vocabulary test at pretest (Time 1), immediate posttest (Time 2), and delayed posttest (Time 3; two weeks later).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendices.docx