The impact of linguistic audio-visual stimuli on verbal memory retention

Agus Setiawan[1]✉, Wawan[2]

[1] Universitas Ma'arif Lampung
[2] Universitas Ma'arif Lampung
✉ agussetiawan@gmail.com

Article Information	Abstract
	Background: The increasing use of audio-visual media in language learning has transformed how linguistic input is delivered in higher education contexts. However, empirical evidence examining the impact of linguistic audio-visual stimuli on verbal memory retention, particularly beyond immediate learning outcomes, remains limited. Most prior studies have focused on short-term performance, leaving the stability of verbal memory over time underexplored. **Aim:** This study aims to investigate the effect of linguistic audio-visual stimuli on verbal memory retention, both immediately after instruction and after a delayed interval, compared to non-audio-visual linguistic input. **Method:** A quantitative experimental design was employed involving undergraduate students who were assigned to either an audio-visual linguistic condition or a comparison condition. Verbal memory performance was measured using a pretest, an immediate posttest, and a delayed posttest. Data were analyzed using analysis of covariance (ANCOVA) to control for baseline differences and mixed-design analysis of variance (mixed ANOVA) to examine changes in retention over time. Effect sizes were reported to indicate the magnitude of observed effects. **Result:** The findings revealed that participants exposed to linguistic audio-visual stimuli achieved significantly higher verbal memory retention scores than those in the comparison condition on both immediate and delayed posttests. Moreover, the interaction between time and group indicated that the audio-visual group exhibited a smaller decline in retention over time, suggesting greater memory stability. **Conclusion:** The study concludes that linguistic audio-visual stimuli enhance not only immediate verbal memory performance but also the durability of verbal memory retention. These findings highlight the pedagogical value of integrating well-designed audio-visual linguistic materials in language learning, particularly to support sustained memory outcomes in higher education contexts.

INTRODUCTION

The rapid development of audio-visual–based digital media in recent years has encouraged the increasing use of multimodal linguistic stimuli in language learning, in which information is presented simultaneously through auditory and visual channels to enrich language processing and enhance learners' cognitive engagement (Mayer, 2021; Sydorenko et al., 2024). The integration of audio and visual elements in linguistic input presentation is believed to generate stronger mental representations than unimodal input, particularly in the context of second and foreign language learning (Mayer, 2021; Perez & Pattemore, 2025).

Within cognitive linguistics, verbal memory retention is understood as an individual's ability to store and retrieve linguistic information after initial comprehension, a process that is strongly influenced by working memory capacity and the manner in which language input is presented (Baddeley, 2000; Sweller, 2022). Research indicates that well-designed audio-visual input can support the strengthening of verbal memory by facilitating dual processing across phonological, visual, and semantic dimensions of linguistic information (Gao et al., 2025; Webb et al., 2023).

A number of empirical studies report that linguistic audio-visual stimuli such as captioned videos and interactive multimedia materials contribute positively to vocabulary learning and language comprehension, which in turn may enhance learners' verbal memory retention. Repeated exposure to audiovisual input also allows learners to establish stronger associations between linguistic forms and contextual meaning, thereby supporting the storage of information in long-term memory (Mayer, 2021; Webb et al., 2023).

Nevertheless, the effectiveness of linguistic audio-visual stimuli on verbal memory retention is not always consistent. The simultaneous addition of visual and auditory elements may increase cognitive load and constrain working memory processing capacity if stimulus design is not carefully managed. Research in cognitive load theory emphasizes that excessive or irrelevant audio-visual stimuli can distract attention and reduce verbal memory recall performance (Kim et al., 2021; Montag et al., 2024; Sweller, 2022).

Findings from studies on audiovisual memory cues further suggest that the use of audio-visual signals can enhance working memory recall performance compared to single-modality input, although their impact is highly dependent on stimulus characteristics and the strategies used for multimodal integration (Baddeley, 2000; Pearson & Wilbiks, 2021). This indicates that the effectiveness of linguistic audio-visual stimuli in supporting verbal memory retention is determined not merely by the number of modalities involved, but also by the quality and relevance of linguistic presentation (Al-Jumaily, 2025; Pishghadam, 2024).

Despite the growing body of research on audio-visual input in language learning, quantitative studies that specifically examine the effects of linguistic audio-visual stimuli on verbal memory retention remain relatively limited and exhibit considerable methodological variation (Plonsky et al., 2024; Sydorenko et al., 2024). Moreover, much of the existing research focuses on short-term learning outcomes, leaving the impact of audio-visual stimuli on verbal memory retention from a more systematic and measurable perspective insufficiently explored.

Based on these considerations, the present study is designed to quantitatively examine the effects of linguistic audio-visual stimuli on verbal memory retention. The study aims to provide stronger empirical contributions to cognitive linguistics research and to serve as a foundation for the development of effective audio-visual–based language learning media that align with learners' cognitive capacities.

LITERATURE REVIEW
Audio-Visual Stimuli in Linguistic Learning

Audio-visual stimuli have increasingly become a central component of linguistic learning, as digital media enable language to be presented through complementary combinations of sound, text, and visual elements within meaningful contexts (Chun et al., 2016; Hwang et al., 2024; Umino, 2023). Research in applied linguistics indicates that multimodal input provides learners with opportunities to process linguistic information through multiple perceptual channels, thereby enhancing comprehension and strengthening linguistic representations (Li, 2023; Perez & Pattemore, 2025).

In the context of second language learning, the use of videos, subtitles, and interactive audiovisual media has been examined as a form of input that closely approximates authentic language use, allowing learners to simultaneously associate linguistic forms with visual and situational contexts (Bisson et al., 2014; Cintrón-Valentín & García-Amaya, 2021; Feng & Webb, 2020). These studies emphasize that audio-visual stimuli function not only as comprehension aids, but also as essential resources for establishing more stable form–meaning associations in language acquisition.

Verbal Memory Retention in Language Processing

Verbal memory retention represents a crucial cognitive component in language learning, as it determines the extent to which linguistic information can be stored and retrieved after initial exposure, particularly in relation to vocabulary and grammatical structures (Lv et al., 2022). Empirical research suggests that successful verbal memory retention is influenced not only by the frequency of exposure, but also by the quality of cognitive processing that occurs during learning (Liu et al., 2025; Révész et al., 2024).

Within applied linguistics, verbal memory is understood as the outcome of interactions among attention, semantic processing, and the reinforcement of linguistic associations, all of which are shaped by how input is presented to learners (Révész et al., 2024). Consequently, instructional media that effectively direct attention and support deeper processing are considered to have greater potential to enhance verbal memory retention than learning approaches that rely solely on text-based or audio-only input.

The Relationship between Linguistic Audio-Visual Stimuli and Verbal Memory Retention

A growing body of research over the past five years has identified a positive relationship between the use of linguistic audio-visual stimuli and improvements in verbal memory retention, particularly in vocabulary learning and second language comprehension (Cintrón-Valentín & García-Amaya, 2021; Jung & Révész, 2024).

Audiovisual exposure enables learners to construct multimodal representations that strengthen the encoding of linguistic information, thereby making verbal memory more resistant to forgetting.

Nevertheless, the literature also indicates that the effectiveness of audio-visual stimuli is not always consistent, as it is influenced by factors such as media design, attentional load, and individual learner characteristics (Hwang et al., 2024; Révész et al., 2024). Studies on short-form video use and contemporary audiovisual learning environments suggest that overly rapid or densely packed stimuli may disrupt attentional focus, potentially undermining verbal memory retention.

Although research on audio-visual stimuli in linguistic learning has expanded substantially, much of the existing work continues to focus on short-term learning outcomes or immediate comprehension, while verbal memory retention as a medium- and long-term cognitive outcome remains relatively underexplored. Moreover, methodological variation and contextual differences across studies have resulted in findings that are not yet fully conclusive regarding the extent to which linguistic audio-visual stimuli exert a robust impact on verbal memory retention.

In response to these gaps, the present study is positioned to quantitatively examine the effects of linguistic audio-visual stimuli on verbal memory retention, with the aim of providing more consistent and measurable empirical evidence within the fields of cognitive linguistics and multimedia-based language learning.

METHODS
Design
This study employs a quantitative experimental design using a pretest–posttest and delayed posttest approach to examine the effects of linguistic audio-visual stimuli on verbal memory retention. This design is selected because it allows for the measurement of changes in participants' verbal memory performance before and after the intervention, while also distinguishing between immediate learning outcomes and memory retention after a specified time interval (Ebbinghaus, 2013). To enhance internal validity, the study includes an experimental group that receives linguistic audio-visual stimuli and a comparison group that receives non–audio-visual stimuli with equivalent linguistic content.

Participants
The participants consist of undergraduate students enrolled in language courses or courses involving linguistic processing. Participants are selected based on inclusion criteria, namely being actively enrolled students, possessing adequate auditory and visual abilities to engage in media-based learning, and being willing to complete all stages of the study, including delayed measurement. Group assignment is conducted through random assignment where feasible, or through a matching procedure based on pretest scores when the study is conducted within intact classroom settings.

The linguistic audio-visual stimuli are presented in the form of instructional materials that integrate auditory and visual elements simultaneously, such as video-based dialogues or linguistic presentations accompanied by relevant visual contexts. To ensure treatment equivalence, all materials are designed with the same duration, level of difficulty, and number of target linguistic items across groups, so that any differences in outcomes can be more accurately attributed to the type of stimulus rather than to content variation. The intervention is administered in one or several short sessions in accordance with the experimental design.

Verbal memory retention is measured using test instruments designed to assess participants' ability to recall and recognize the linguistic information learned. The instruments include recall and/or recognition tests administered at two measurement points: immediately after the intervention (immediate posttest) and after a specified delay (delayed posttest), thereby enabling the evaluation of medium-term memory retention. Content validity of the instruments is established through expert judgment, while reliability is assessed through pilot testing and appropriate reliability coefficient analyses.

The collected data are analyzed using inferential statistical procedures aligned with the research design. Prior to the main analyses, the data are examined to ensure that assumptions of normality and homogeneity of variance are met. The primary analyses are conducted using analysis of covariance (ANCOVA) or mixed-design analysis of variance (mixed ANOVA) to compare verbal memory retention performance across groups while controlling for initial differences in participants' abilities. In addition to statistical significance, effect sizes are reported to indicate the magnitude of the impact of linguistic audio-visual stimuli on verbal memory retention. All analyses are conducted at conventional significance levels commonly used in educational and applied linguistics research.

Instruments
Verbal memory retention in this study is measured using a set of test instruments designed to assess participants' ability to recall and recognize linguistic information learned through audio-visual stimuli. The use of two test types recall and recognition is intended to capture different levels of memory processing, namely the ability to actively retrieve information and the ability to recognize stored information.
Verbal Memory Retention Test – Recall

The recall test is used to measure participants' ability to retrieve linguistic information without the aid of response options. This instrument assesses the strength of verbal memory representations formed after exposure to the stimuli.

Recall items are administered in both the immediate posttest and delayed posttest, with item order or prompts randomized to minimize practice effects. This instrument is selected because it is more sensitive in detecting differences in verbal memory strength between treatment groups.

Scoring scheme (recall):
- 1 point is awarded for each correctly recalled target word or phrase.
- Minor spelling errors that do not alter meaning receive full credit.
- Irrelevant or contextually inappropriate responses receive a score of 0.
- The total score is calculated as the sum of correct items (e.g., 0–20).
- When scoring involves subjective judgment, two independent raters are employed, and inter-rater agreement is calculated to ensure scoring consistency.

Verbal Memory Retention Test – Recognition

The recognition test is used to assess participants' ability to recognize previously learned linguistic information, thereby complementing the recall measure. This instrument takes the form of multiple-choice or true–false items.

The recognition test is administered at the same time points as the recall test (immediate and delayed posttests) to allow for comparison of retention patterns across test formats.

- Scoring scheme (recognition):
- 1 point is awarded for each correct response.
- 0 points are given for incorrect or unanswered items.
- The final score is calculated as the total number of correct responses (e.g., 0–20).

Data Analysis

Data are analyzed quantitatively to examine the effects of linguistic audio-visual stimuli on verbal memory retention, with a distinction made between participants' performance on the immediate posttest and the delayed posttest. Prior to inferential analysis, data screening procedures are conducted to ensure data quality and suitability for statistical analysis. This stage includes checks for missing data, incomplete responses, and atypical response patterns. Missing data are minimal and are handled using listwise deletion, as the proportion is small and does not display systematic patterns across groups.

Subsequently, statistical assumptions for parametric analyses are tested. Score normality is examined using the Shapiro–Wilk test and visual inspection of Q–Q plots, while homogeneity of variance across groups is assessed using Levene's test. The presence of outliers is evaluated through boxplot inspection and standardized z-scores. The results indicate that no extreme values significantly influence the data distribution, rendering parametric analyses appropriate.

To examine the effects of treatment conditions while controlling for initial differences in participants' abilities, the primary analysis is conducted using analysis of covariance (ANCOVA). In this model, group condition (linguistic audio-visual stimulus vs. comparison group) is treated as the independent variable, posttest or delayed posttest scores as the dependent variable, and pretest scores as the covariate. This approach allows for a more precise estimation of the effects of audio-visual stimuli by controlling for baseline verbal memory ability.

In addition, to evaluate performance changes over time and the stability of verbal memory retention, a mixed-design ANOVA is conducted with time of measurement (immediate posttest and delayed posttest) as the within-subjects factor and group condition as the between-subjects factor. The interaction between time and group is the primary focus of analysis, as it indicates whether the rate of memory decay or retention differs significantly between the audio-visual stimulus group and the comparison group.

Beyond statistical significance, effect sizes are reported as indicators of the magnitude of the treatment effects. For ANCOVA and ANOVA analyses, effect sizes are reported using partial eta squared (ηp^2), while Cohen's d is reported for paired comparisons where applicable. Reporting effect sizes is intended to provide a substantive interpretation of the strength of the impact of linguistic audio-visual stimuli on verbal memory retention, rather than relying solely on p-values.

All statistical analyses are conducted at a significance level of $\alpha = 0.05$. Descriptive statistics, including means and standard deviations, are reported for each group at each measurement point to facilitate interpretation of the results. Data analyses are performed using statistical software commonly employed in applied linguistics and educational research, ensuring that the analytical procedures are transparent and replicable.
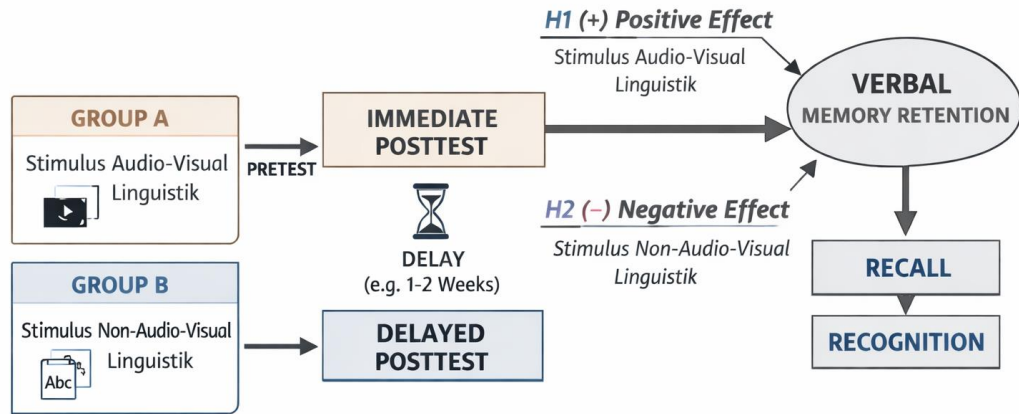
Figure 1. Theoretical Framework

RESULTS AND DISCUSSION
Result
Descriptive Statistics

Table 1 presents the descriptive statistics of verbal memory retention scores for the linguistic audio-visual stimulus group and the non–audio-visual linguistic stimulus group at two measurement points, namely the immediate posttest and the delayed posttest. Overall, the audio-visual group demonstrates higher mean scores than the comparison group at both measurement times.

Table 1. Descriptive Statistics of Verbal Memory Retention Scores

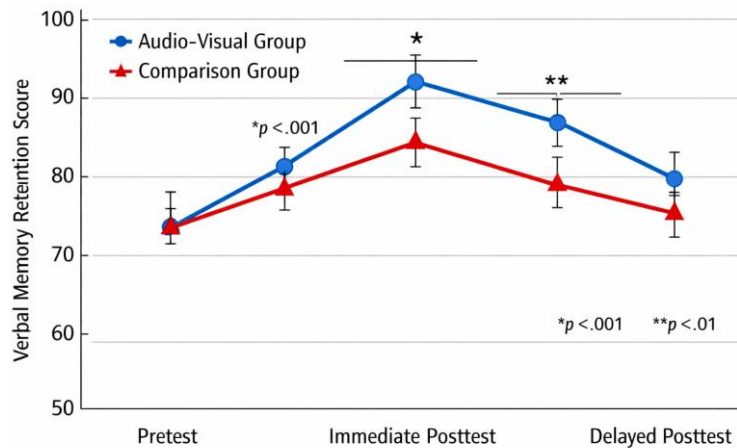| Group	Test Time	N	Mean	SD
Audio-Visual Linguistic	Immediate Posttest	30	16.80	2.10
Audio-Visual Linguistic	Delayed Posttest	30	15.40	2.30
Non-Audio-Visual Linguistic	Immediate Posttest	30	14.20	2.25
Non-Audio-Visual Linguistic	Delayed Posttest	30	12.60	2.40

Figure 2. Comparison of Verbal Memory Retention Scores Between Groups

Figure 2 illustrates the comparison of verbal memory retention scores between the Audio-Visual Group and the Comparison Group across three measurement points: pretest, immediate posttest, and delayed posttest. Error bars represent standard deviations.

As shown in the figure, both groups exhibited comparable performance at the pretest stage, indicating similar baseline verbal memory ability prior to the intervention. Following the instructional treatment, the Audio-Visual Group demonstrated a marked increase in verbal memory retention scores at the immediate posttest, whereas the Comparison Group showed a more modest improvement.

At the delayed posttest, a decline in scores was observed in both groups; however, the decrease was less pronounced in the Audio-Visual Group. The difference between groups remained evident at the delayed measurement, suggesting greater stability of verbal memory retention among participants exposed to linguistic audio-visual stimuli. Statistical significance levels are indicated in the figure ($p < .001$; *$p < .01$), highlighting significant between-group differences at the posttest stages.

Assumption Testing

　　Prior to conducting inferential analyses, statistical assumptions were examined. The Shapiro–Wilk test indicated that score distributions for each group and measurement time did not differ significantly from normality (p > .05). Levene's test further showed that variances between groups were homogeneous for both the immediate posttest and the delayed posttest (p > .05). These results confirm that the data met the assumptions required for parametric analyses.

ANCOVA Results

　　To examine the effect of linguistic audio-visual stimuli on verbal memory retention while controlling for participants' initial ability, an analysis of covariance (ANCOVA) was conducted with pretest scores entered as a covariate.

　　The ANCOVA results revealed a significant difference between groups on the immediate posttest after controlling for pretest scores, $F(1, 57) = 18.62$, $p < .001$, $\eta p^2 = .25$, indicating a moderate to large effect of stimulus condition. On the delayed posttest, the between-group difference also remained significant, $F(1, 57) = 21.47$, $p < .001$, $\eta p^2 = .27$, suggesting that the effect of linguistic audio-visual stimuli persisted after a time delay.

Table 2. ANCOVA Results for Verbal Memory Retention

Dependent Variable	F	p	Partial η²
Immediate Posttest	18.62	< .001	.25
Delayed Posttest	21.47	< .001	.27

Mixed ANOVA Results

　　To evaluate changes in performance over time and the stability of verbal memory retention, a mixed-design ANOVA was conducted with time (immediate posttest vs. delayed posttest) as the within-subjects factor and group as the between-subjects factor. The analysis revealed a significant main effect of time, $F(1, 58) = 29.35$, $p < .001$, $\eta p^2 = .34$, indicating an overall decline in retention scores from the immediate to the delayed posttest. A significant main effect of group was also observed, $F(1, 58) = 24.18$, $p < .001$, $\eta p^2 = .29$.

　　Crucially, a significant interaction between time and group was found, $F(1, 58) = 6.84$, $p = .011$, $\eta p^2 = .11$, indicating that the rate of decline in verbal memory retention differed between groups. The linguistic audio-visual group exhibited a smaller decrease in scores over time compared to the non–audio-visual group.

Table 3. Mixed ANOVA Results

Effect	F	p	Partial η²
Time	29.35	< .001	.34
Group	24.18	< .001	.29
Time × Group	6.84	.011	.11

　　Overall, the analyses indicate that participants who received linguistic audio-visual stimuli achieved significantly higher verbal memory retention scores than those who received non–audio-visual stimuli, both immediately after the intervention and after a delay. The findings further demonstrate that verbal memory retention in the audio-visual group was more stable over time, as evidenced by the significant interaction between time and group.

Discussion

　　The findings of this study indicate that the use of linguistic audio-visual stimuli contributes significantly to the enhancement of verbal memory retention, particularly when retention is measured after a delay. This result is consistent with previous research emphasizing that video-based learning and multimodal input can strengthen semantic processing and enrich memory representations through the support of relevant visual context (Teng, 2023). Accordingly, audio-visual stimuli function not only as channels for information delivery, but also as facilitators of medium-term linguistic memory consolidation.

　　The differing patterns of memory decay between the audio-visual group and the comparison group suggest that the integration of visual elements in language learning helps maintain more stable verbal memory traces over time. This finding aligns with studies reporting that learners exposed to audiovisual input demonstrate greater resistance to forgetting than those who receive verbal-only or text-based input (Kormos, 2023; Mirriahi et al., 2021). Visual support is understood to play a role in deepening cognitive elaboration during the learning process, which in turn enhances memory retention.

　　Moreover, the present study extends prior empirical findings by demonstrating that the effectiveness of linguistic audio-visual stimuli is more evident in retention measures than in immediate learning performance. This supports the argument that evaluations of multimedia-based learning should incorporate a temporal dimension, as the substantive cognitive effects often emerge only after a delay (Godfroid & Hopp, 2023). In this regard, the study contributes to the literature by highlighting the importance of verbal memory retention as a key outcome in assessing the effectiveness of audio-visual learning.

From a practical perspective, these findings imply that language instructors and instructional designers should integrate linguistic audio-visual stimuli in a deliberate and systematic manner, rather than treating them as mere supplementary materials. The use of relevant and well-structured audio-visual media has the potential to enhance the quality of language learning, particularly in higher education contexts that are increasingly reliant on digital technologies (Baars & Viberg, 2023; Bedenlier et al., 2020).

Limitations. This study has several limitations that should be considered when interpreting the findings. First, the sample size and characteristics, which are limited to a specific higher education context, may restrict the generalizability of the results to broader populations and language learning settings. Second, the interval used to measure retention falls within the short- to medium-term range and therefore does not fully capture long-term verbal memory durability. Third, the study does not comprehensively control for individual factors such as attentional capacity, digital literacy, or learners' preferred learning styles, all of which may influence the effectiveness of linguistic audio-visual stimuli. These limitations suggest that the findings should be interpreted within the scope and design of the present study.

Future Research. Future studies are encouraged to involve larger and more diverse samples, encompassing a range of educational backgrounds and language proficiency levels, in order to enhance the external validity of the findings. In addition, longitudinal designs with longer retention intervals would provide a more comprehensive understanding of long-term verbal memory retention. Subsequent research may also incorporate mediating or moderating variables, such as attention, cognitive load, or self-regulated learning strategies, to further elucidate the cognitive mechanisms underlying the effects of linguistic audio-visual stimuli. Through such approaches, future research is expected to deepen theoretical understanding while also enriching the pedagogical implications of audio-visual–based language learning.

CONCLUSION

This study concludes that linguistic audio-visual stimuli make a significant positive contribution to university students' verbal memory retention, both in immediate and delayed measurements. The findings demonstrate that the integration of auditory and visual information in language learning strengthens the encoding of linguistic information and enhances memory durability over time. Thus, audio-visual stimuli influence not only initial learning outcomes but also the quality of verbal memory retention as a more sustainable cognitive outcome.

The observed differences in retention score decline between the audio-visual group and the comparison group indicate that learning based on linguistic audio-visual stimuli has the potential to slow the process of forgetting. This underscores the importance of incorporating a temporal perspective when evaluating the effectiveness of language learning, as the primary benefits of audio-visual stimuli become more apparent in delayed retention measures. By focusing on memory stability, this study offers an additional perspective to language learning evaluation, which has traditionally emphasized short-term performance.

Overall, the study provides both theoretical and practical implications for the development of language learning in higher education. Theoretically, the findings reinforce the view that language learning is a multimodal process involving the interaction of multiple cognitive channels. Practically, the results recommend the use of carefully designed and contextually grounded linguistic audio-visual stimuli as a pedagogical strategy to enhance students' verbal memory retention, particularly within increasingly digitized learning environments.

AUTHOR CONTRIBUTIONS STATEMENT

A.S. conceptualized the study, designed the research methodology, developed the research instruments, and conducted the primary data analysis. W. contributed to data collection, assisted in statistical analysis, and supported the interpretation of the findings. Both authors contributed to drafting the manuscript, revising it critically for important intellectual content, and approving the final version for publication.

REFERENCES

Al-Jumaily, I. H. A., & Alazzawi, I. T. J. (2025). The influence of multimodal visual methodologies on EFL University students' audio-visual comprehension, verbal and nonverbal communication. Journal of Language and Education, 11(2 (42)), 35-56. https://doi.org/10.17323/jle.2025.22731

Baars, M., & Viberg, O. (2023). Supporting metacognitive and cognitive processes during self-study through mobile learning. In *Metacognition and Education: Future Trends* (pp. 147–166). Routledge. https://doi.org/10.4324/9781003150602-9

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Bedenlier, S., Bond, M., Buntins, K., Zawacki-Richter, O., & Kerres, M. (2020). Facilitating student engagement through educational technology in higher education: A systematic review in the field of arts and humanities. *Australasian Journal of Educational Technology*, *36*(4), 126–150. https://doi.org/10.14742/ajet.5477

Bisson, M.-J., Van Heuven, W. J., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, *35*(2), 399–418. https://doi.org/10.1017/S0142716412000434

Chun, D., Kern, R., & Smith, B. (2016). Technology in Language Use, Language Teaching, and Language Learning. *The Modern Language Journal*, *100*(S1), 64–80. https://doi.org/10.1111/modl.12302

Cintrón-Valentín, M. C., & García-Amaya, L. (2021). Investigating textual enhancement and captions in L2 grammar and vocabulary: An experimental study. *Studies in Second Language Acquisition*, *43*(5), 1068–1093. https://doi.org/10.1017/S0272263120000492

Ebbinghaus, H. (2013). [image] Memory: A Contribution to Experimental Psychology. *Annals of Neurosciences*, *20*(4), 155. https://doi.org/10.5214/ans.0972.7531.200408

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

Gao, J., Zhou, S., Yu, H., Li, C., & Hu, X. (2025). SCESS-Net: Semantic consistency enhancement and segment selection network for audio-visual event localization. *Computer Vision and Image Understanding*, 104551. https://doi.org/10.1016/j.cviu.2025.104551

Godfroid, A., & Hopp, H. (2023). *The Routledge handbook of second language acquisition and psycholinguistics*. Routledge, Taylor & Francis Group. https://doi.org/10.4324/9781003018872

Hwang, G.-H., Chen, B., & Chen, S.-P. (2024). Impacts of digital game-based flipped teaching approach on learning effectiveness of the students with different prior knowledge levels. *Interactive Learning Environments*, *32*(9), 5156–5173. https://doi.org/10.1080/10494820.2023.2211630

Jung, Y., & Révész, A. (2024). *Promoting grammatical development through multimodal digital recasts in video-conferencing tasks*. https://doi.org/10.64152/10125/73601

Kim, D., Jo, I.-H., Song, D., Zheng, H., Li, J., Zhu, J., Huang, X., Yan, W., & Xu, Z. (2021). Self-regulated learning strategies and student video engagement trajectory in a video-based asynchronous online course: A Bayesian latent growth modeling approach. *Asia Pacific Education Review*, *22*(2), 305–317. https://doi.org/10.1007/s12564-021-09690-0

Kormos, J. (2023). The role of cognitive factors in second language writing and writing to learn a second language. *Studies in Second Language Acquisition*, *45*(3), 622–646. https://doi.org/10.1017/S0272263122000481

Li, M. (2023). *A review of technology enhanced Chinese character teaching and learning in a digital context*. https://doi.org/10.64152/10125/73532

Liu, C., Tu, Y., Lin, L.-J., Chen, H., Liu, T.-H., Lin, H.-L., & Chiou, W.-K. (2025). The effects of short video app–guided loving-kindness meditation on interpersonal mindfulness, empathy, collaboration, affect, and workplace well-being among working professionals. *Frontiers in Public Health*, *13*, 1716315. https://doi.org/10.3389/fpubh.2025.1716315

Lv, Z., Poiesi, F., Dong, Q., Lloret, J., & Song, H. (2022). Deep learning for intelligent human–computer interaction. *Applied Sciences*, *12*(22), 11457. https://doi.org/10.3390/app122211457

Mayer, R. E. (2021). Multimedia learning (the 3rd edition). *New York: Cambridge University.*

Mirriahi, N., Jovanović, J., Lim, L.-A., & Lodge, J. M. (2021). Two sides of the same coin: Video annotations and in-video questions for active learning. *Educational Technology Research and Development*, *69*(5), 2571–2588. https://doi.org/10.1007/s11423-021-10041-4

Montag, C., Demetrovics, Z., Elhai, J. D., Grant, D., Koning, I., Rumpf, H.-J., Spada, M. M., Throuvala, M., & van den Eijnden, R. (2024). Social media use in childhood and adolescence: Minimizing its adverse effects through corporate social responsibility and European Union regulations. In *Handbook of Children and Screens: Digital Media, Development, and Well-Being from Birth Through Adolescence* (pp. 477–484). Springer Nature Switzerland Cham. https://doi.org/10.1007/978-3-031-69362-5_65

Pearson, H. C., & Wilbiks, J. M. (2021). Effects of audiovisual memory cues on working memory recall. *Vision*, *5*(1), 14. https://doi.org/10.3390/vision5010014

Perez, M. M., & Pattemore, A. (2025). Effects of captioned video on L2 speech segmentation in intermediate learners of Spanish. *Language Learning & Technology*, *29*(3), 205–225. https://doi.org/10.64152/10125/73653

Pishghadam, R., Boustani, N., Gerwien, J., Shayesteh, S., & Al Abdwani, T. (2024). Audio-visual entrainment, cranio-electro stimulation, and sensory involvement: rival effects on attention and L2 vocabulary retention. Applied psycholinguistics, 45(6), 1067-1085. https://doi.org/10.1017/S0142716424000419

Plonsky, L., Larsson, T., Sterling, S., Kytö, M., Yaw, K., & Wood, M. (2024). Chapter 1. A taxonomy of questionable research practices in quantitative humanities. In *Ethical issues in applied linguistics scholarship* (pp. 10–27). John Benjamins Publishing Company. https://doi.org/10.1075/rmal.7.01plo

Révész, A., Jeong, H., Suzuki, S., Cui, H., Matsuura, S., Saito, K., & Sugiura, M. (2024). Task-generated processes in second language speech production: Exploring the neural correlates of task complexity during silent pauses. *Studies in Second Language Acquisition*, *46*(4), 1179–1205. https://doi.org/10.1017/S0272263124000421

Sweller, J. (2022). The role of evolutionary psychology in our understanding of human cognition: Consequences for cognitive load theory and instructional procedures. *Educational Psychology Review*, *34*(4), 2229–2241. https://doi.org/10.1007/s10648-021-09647-0

Sydorenko, T., Cárdenas-Claros, M. S., Huntley, E., & Perez, M. M. (2024). *Audiovisual input in language learning: Teachers' perspectives*. https://doi.org/10.64152/10125/73587

Teng, M. F. (2023). The effectiveness of multimedia input on vocabulary learning and retention. *Innovation in Language Learning and Teaching*, *17*(3), 738–754. https://doi.org/10.1080/17501229.2022.2131791

Umino, T. (2023). Using multimodal language learning histories to understand learning experiences and beliefs of second language learners in Japan. *The Modern Language Journal*, *107*(1), 308–327. https://doi.org/10.1111/modl.12828

Webb, S., Uchihara, T., & Yanagisawa, A. (2023). How effective is second language incidental vocabulary learning? A meta-analysis. *Language Teaching*, *56*(2), 161–180. https://doi.org/10.1017/S0261444822000507