The Impact of Teachers' Multimodal Cues on Students' L2 Vocabulary Learning in Naturalistic Classroom Teaching

Jing Zhou (amy.zhou@autuni.ac.nz)
Department of Psychology, Auckland University of Technology, NZ
Department of Foreign Languages, Chaohu University, CH

Yan Gu (yan.gu@essex.ac.uk)
Department of Psychology, University of Essex, UK
Department of Experimental Psychology, UCL, UK

Abstract

We investigated the impact of teachers' multimodal cues on L2 word learning in naturalistic teaching. 169 university students randomly watched 12 of 54 clips of English vocabulary instructions and took subsequent word recognition and learning tests. The learning outcomes were analysed as a function of teachers' prosodic, linguistic and gestural input during the instruction of each vocabulary while controlling for students' characteristics and varying teachers' influences. Results showed that a shorter mean length of utterances, fewer L2 English words, and more questions for students and "phrase" teaching predicted better learning outcomes. Furthermore, students learning improved with teachers' slower speaking rate but fewer pauses and more iconic gestures. These results were robust even after controlling for other significant factors such as students' English proficiency, working memory, degree of liking of teachers and different teachers. Overall, multimodal cues enhance L2 vocabulary learning, with implications for educators, linguists, and cognitive scientists.

Keywords: multimodal cues; vocabulary learning; classroom instruction; MLU; questions; gestures; speaking rate

Introduction

In second language (L2) teaching, multimodal cues such as linguistic features, speech prosody and gestures become integral. These cues, extending beyond verbal interaction, facilitate comprehension (e.g., Takahashi et al., 2018) and enhance instructional appeal and effectiveness (e.g., Lajevardi et al., 2017). To process multimodal cues, one needs to intuitively decode language structures from complex stimuli without explicit instruction (Aslin & Newport, 2012; Lee, 2023). This process primarily involves automatically extracting statistical regularities and form-function pairings from linguistic inputs through passive exposure (Moser et al., 2021). This is often called statistical learning, fundamental in L1 and L2 language acquisition.

Building on this foundation, multimedia-enhanced language learning suggests that L2 learners benefit from receiving multiple sensory cues, such as auditory, visual, textual, and kinesthetic inputs, which enhance cognitive processing and improve learning outcomes (Pujadas & Muñoz, 2020). Mirzaei et al. (2023) further found that multimodal inputs combining auditory, visual, and textual cues are more effective than unimodal inputs in enhancing the retention and recall of vocabulary sets. While these studies show that varied inputs can affect learning, they often focus on limited cues without establishing the relative importance or considering students' characteristics. In this study, we comprehensively analysed the effects of multimodal teaching on L2 word learning.

Multimodal cues in L2 learning

Previous studies have shown that linguistic contents, prosody and gestures can influence student learning outcomes separately. First, the intricate dynamics of linguistic features such as Mean Length of Utterance (MLU), the label of naming or repetition, questioning techniques, code-switching and the use of phrases in language teaching all independently play a pivotal role in influencing students' L2 learning. For example, a longer MLU is linked to richer vocabularies in children (Hart & Risley, 1996). In contrast, for learners of English as an *Additional Language*, vocabulary growth is positively associated with the total number of words used by the teacher but negatively related to MLU (Bowers & Vasilyeva, 2011). Furthermore, the benefits of repetition for word learning have been acknowledged (Berthier & Lambon, 2014). The role of repetition is presumed to create a sensorimotor representation of the word in the learners' minds, allowing the learners to remember this word better (Krishnan et al., 2017). However, over-reliance on repetition may lead to superficial learning, where learners memorise without fully understanding the content (Larsen-Freeman & Anderson, 2013). As for questioning techniques, they are crucial for promoting children's reasoning, math, and verbal explanations (Duong et al., 2021; Tompkins et al., 2017). For example, caregivers' Yes/No questions facilitate children's both immediate and long-term word learning (Dong et al., 2021). In addition, the interplay between the L1 and L2 in teaching settings is crucial. While L1 can aid in understanding vocabulary through direct translations (Rauf, 2018), excessive L2 use may overwhelm learners, whereas insufficient use may hinder language immersion (Edstrom, 2009). Furthermore, phrases are composed of words familiar to learners, suggesting that a phrase might facilitate easier recognition and learning (Smith & Murphy, 2015). Nevertheless, these studies have analysed different linguistic cues separately, primarily centring on text-based speech interactions.

Second, prosodic features such as pitch and speaking rate have been extensively studied for their impact on language

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

learning outcomes. In child-directed language, the degree of pitch modification by mothers when introducing unfamiliar words predicts their children' s immediate word recognition and vocabulary growth (Shi et al., 2023). Similarly, pitch enhancement can facilitate adult vocabulary learning across different visual contexts (Filippi et al., 2014). Adjusting the pitch range can subtly guide secondary school learners without explicit or direct feedback (Sikveland et al., 2021). However, pitch becomes less salient in adult speech, and its impact on L2 vocabulary learning is understudied. As for speaking rate, the focus has primarily been on child-directed speech. For example, a slower speaking rate significantly enhances infants' ability to recognize words (Song et al., 2010) and can facilitate language learning even before children start speaking (Raneri et al., 2020). The few existing studies on adults showed that speech rate establishes listener expectations, influencing their sensitivity to the overall speech pace and subsequently affecting word recognition. These influences on perception and comprehension have intensified over time (Baese-Berk et al., 2014). However, Munro and Derwing (1998) found that while native listeners may prefer slower rates of accented speech, simply slowing down may not benefit L2 learners. Thus, the role of speaking rate in L2 word learning is still unclear, let alone its interaction with other cues.

Third, gestures significantly enhance language teaching and learning by improving students' memory and learning efficiency (Stam et al., 2012; Stam & Tellier, 2022). Tellier (2008) demonstrated that iconic gestures used by teachers and mimicked by students markedly boost memorization of L2 words. Similarly, learners who repeated gestures associated with new words showed a marked improvement in memory retention (Cook et al., 2008). This suggests that gestures serve as visual and motor modalities in deepening the memorization of L2 vocabulary. Further, gestures help students remember words, especially when they naturally suggest the word or its pronunciation (Clark & Trofimovich, 2016). Wang et al. (2023) also found that students who received gesture training outperformed the control group in overall academic presentation performance, underscoring the role of gestures in enhancing expressive abilities. However, gestures are not always beneficial. Kelly and Lee (2012) observed that gestures play different roles since they aid in learning simple pairs but hinder complex ones, implying that the effectiveness of gesture integration depends on the phonetic complexity of the learning material. Moreover, frequent use of naturally modelled beat gestures can increase cognitive load and impair language comprehension, thus advocating a limited use of beat gestures in extended discourse for language learners (Rohrer et al., 2020). These findings highlight the need for further exploration to delineate how different types of gestures affect language learning when accounting for factors like the learner's language proficiency, working memory, and the characteristics of the speech input, including prosody and linguistic features.

While several studies have addressed how a combination of two or three cues contributes to language comprehension and learning, there is a scarcity of research quantifying the impact of each cue when multimodal cues are considered together in L2 instruction. For instance, Drijvers and Holler (2023) highlighted a multimodal facilitation effect, illustrating that participants understood spoken words faster when combined with visual input. Similarly, using prosodic and gestural prominence together was highlighted as potentially constituting a good teaching strategy in L2 teaching (Kushch et al., 2016, 2018). However, the specific impact of these combined cues on learning outcomes has yet to be explored. Furthermore, Donnellan et al. (2023) provided evidence for the significance of multimodal caregiver behaviors, especially prosody, in children's lexical development. However, child-directed language has salient prosody and a higher gesture rate but shorter MLU and simpler lexicons. It is unknown how reallocating the communication weight of different modalities will reshuffle the relative importance of each cue in adult-directed language. Zhang et al. (2023) provided in-depth insights into how prosody, gestures and mouth movements affect brain activity during L2 comprehension in naturalistic contexts, emphasizing the varied impact of these cues on L2 learners compared to native speakers. Nevertheless, they did not investigate the impact of cues on learning.

Internal factors in L2 learning

Students' internal factors, such as working memory, L2 English proficiency and degree of liking teachers may affect their learning outcomes. Better working memory allows for more effective focus and information encoding, thus enhancing students' learning efficiency (Cowan, 2012; Linck et al., 2014). Higher L2 proficiency is directly linked to improved reading performance and broader learning capabilities, including vocabulary acquisition (Jiang, 2011). It is also associated with greater integrative motivation among learners, leading to active engagement with the language (Samad et al., 2012). This fosters a positive learning environment, enhances learner confidence and promotes participation in discussions and activities, which are pivotal for successful language acquisition and improved learning outcomes (Martirosyan et al., 2014). Furthermore, students' liking for their teachers leads to more effort, concentration, and persistence in classes (Saito et al., 2018), which fosters greater engagement in learning (Fredrick, 1980). Moreover, a supportive teacher-student relationship impacts behavioural and instructional engagement and is a key predictor of learning performance (Baafi, 2020).

The current study

While communication is multimodal, we still have an incomplete understanding of the impact of multimodal cues (linguistic features, prosody, and gestures) and their interplay in a teaching context. Furthermore, most studies focus on L1, which may not be applied to an L2 context with diverse processing challenges. Therefore, our study examined the

joint impact of multimodal cues in naturalistic L2 teaching and quantified the respective role and weight of each cue while considering individual differences.

Gestures and prosody have reduced cognitive load in word processing (Osorio et al., 2023; Zhang et al., 2021). We anticipate teachers using slower speech, shorter utterances, and meaningful gestures to enhance learning effectiveness. Furthermore, factors such as students' processing capabilities (working memory), language proficiency (L2 skills), and preferences for different teachers will influence learning outcomes.

Method

Participants

275 Chinese students (M=20.08 years, SD=1.49, 164 men;111 women) from Chaohu University in China were invited to participate in an online study as volunteers. They primarily majored in Mechanical Engineering, Business Administration, and the Arts. Participants signed online consent forms for the study, and Chaohu University approved the research.

Materials

We extracted 54 video clips of vocabulary instruction (M=26.01sec, SD=19.82) from 14 classroom recordings of 4 EFL (English as a foreign language) teachers. The clips varied in speaking rate (M=3.71 syllables/sec, SD=0.76) and mean pitch (M=239.9 Hz, SD=30.93), MLU (M=7.13 words, SD=4.44), and the total number of English words (M=16.2, SD=12.28). Teachers granted consent to use their videos.

In all clips, educators provided clear definitions and explanations of the vocabulary, employing a blend of Chinese and English teaching methods. These segments also included a variety of nonverbal cues to ensure a thorough representation of diverse multimodal cues. Preliminary findings from our pilot study indicated that viewers experienced fatigue after viewing 12 videos. Consequently, we developed six different versions of the stimulus material, each consisting of 12 clips. Certain clips were consistently used across different versions.

To mitigate the impact of cognitive and memory variations on students' learning outcomes, we adapted a working memory test from the Arealme website (*https://www.arealme.com/memory-test/cn/).* It had 13 items, including 9 textual, 3 visual images, and 1 numerical information.

Procedure

Rating of liking teachers Participants were first randomly assigned to watch 12 clips of L2 English vocabulary instruction from one of the six versions and rated the extent to which they liked the teaching (scale 1-10) in each clip via a Tencent questionnaire. Students' demographic data were collected, including gender, age, major and self-evaluated English proficiency (scale 1-10, M=6.26, SD=2.0), etc.

Word recognition test After watching all the video clips, students answered 15 questions (including 12 words that they had watched the video explanations and 3 distractors). For each vocabulary, students were presented with four options:

Option A: I have not been taught this word.

Option B: I am unsure if I have been taught this word.

Option C: I have seen the video explanation of this word, but I do not remember its meaning.

Option D: I have seen the video explanation of this word, and I know its meaning.

Inspired by Montero et al. (2014), we developed a vocabulary recognition test to measure vocabulary acquisition. By distinguishing 'known' from 'unknown' words (Hulstijn, 2001), this method offered a clearer view of students' vocabulary progress and provided feedback on their learning. In the assessment, if students chose Option D, they were asked if they knew the word before or after watching the video. We determined success in word acquisition if the student recognized the word after watching the video and correctly identified its meaning in a four-choice question.

Attentional questions and working memory test Students completed five logic puzzles to assess their attention and took a ten-question working memory test.

Coding and Measures

Teacher's multimodal input

Linguistic features: Speech transcriptions: the first author transcribed teachers' utterances, which were rigorously reviewed and cross-checked against the videos for precision. For transcriptions of each video, we measured (1) MLU, calculated as the average number of words per utterance (Sun & Verspoor, 2022). (2) *Total questions*: the number of questions the teacher asked per the video transcript. (3) *Total number of words and number of English words*: the number of words in a clip. Since teachers used English and Chinese during vocabulary instruction, we separately calculated the number of English syllables and Chinese characters. (4) *Label of Naming*: the frequency of mentions for each target word during its explanation (Candry et al., 2018). (5) *Phrases*: the two researchers categorized an instructed vocabulary into an individual word or phrase (such as "throw-away").

Prosody: Boundaries of utterances were annotated in Praat (Boersma & Weenink, 2023). A Praat script extracted each vocabulary instruction's utterance duration, pitch, and intensity values. We computed the following measures: (1) *Speaking rate*: the average number of syllables per second excluding pauses over 200 ms (Han, 2019). (2) *Pitch*: mean F0 and F0 range, transformed to semitones (Shi et al., 2023). (3) *Intensity*: mean intensity and intensity range. (4) *Pauses*: mean pausing duration and pausing rate.

Non-verbal cues: *Gestures*: the first author coded the gestures in all clips using ELAN, and a second coder verified this coding. Intercoder reliability was high, with 94.44% agreement and a Kappa coefficient of 0.88, indicating substantial consistency. Discrepancies were resolved through

discussions between the coders: (1) Iconic gestures: relating to semantic meaning, e.g., a kicking motion with the hand to depict the action of 'kick' (McNeill, 1992). (2) Beat gestures: a hand moving up and down to make an emphasis, aligning with the prosodic rhythm of speech. (3) Pointing: referring to concrete events (pointing to a pen) or abstract ideas (e.g., space and time) (Hudson, 2011). (4) Interactive gestures: moving hands in a circling way unconsciously to maintain the process of the dialogue (Bavelas et al., 1995). The gesture rate was computed by the number of gestures per 100 words. *Gaze*: During the vocabulary instruction, some teachers made eye contact with students, while others mostly looked at the textbook or PowerPoint and did not look at students. Eye gaze, by focusing learners' attention on key materials (Kuang et al., 2023) and enhancing the connection between learners and instructors, significantly improves learning outcomes and student engagement (Sharma et al., 2016). We coded each video whether the teacher gazed at students or not.

Coding of student learning outcome

We categorized the learning outcomes for unknown words. Correct answers were coded as '1', while incorrect answers and responses indicating that the participant did not know a word's meaning were both coded as '0'. Target words already known to a participant before the vocabulary instruction were excluded from the analysis.

Data Analysis

To ensure data integrity, participants who did not complete the survey (N=106) or provided uniform responses across the entire questionnaire (N=13) were excluded. In addition, those who had a score of 0 for working memory or attentional test items were also excluded (N=12). Thus, the data of the remaining 144 participants were submitted for analysis.

Generalized linear mixed-effects models using the glmmTMB package in the R environment (R core team, 2020) were used to assess students' vocabulary learning outcomes (DV: binary for each word). Prosodic, linguistic and gestural cues were analyzed as main predictors, with control variables of different teachers, students' working memory, English proficiency, degree of liking the teachers, and versions of stimuli. We added a by-participant and by-item random intercept. No random slope was added as it would lead to an error message of model convergence. As MLU is often highly related to speaking rate (Malécot et al., 1972), we ran a separate regression either with MLU or speaking rate independently. We used the AIC values to compare and select models.

Result

Table 1 presents descriptive information about some main variables and their corresponding average learning outcomes. We started with a full model of all predictors. The initial results showed that teachers' speaking rate, pausing rate, use of questions, English words, teaching phrases, producing a gesture, and students' English proficiency, working memory, and degree of liking teachers were significant predictors for student learning outcomes, whereas teachers' mean pitch, MLU, number of mentions of the label, whether gazing at students, total speaking duration of a vocabulary was not significant. There was no multicollinearity between variables in the model (all VIFs < 3). After various model comparisons, the final best-fit model showed that the effects mentioned above were robust, except that MLU was also significant (β = -.08, p = .011) in its model, excluding speaking rate (as MLU indeed strongly correlated with the speaking rate (r = 0.57, p < .001)). Details of the result are as follows.

Table 1: Overview of key predictors for learning outcomes.

IVs	Features	Group	Mean	SD
Linguistic features	MLU	Higher	25.51%	0.44
		Lower	29.28%	0.46
	Questions	Higher	29.91%	0.46
		Lower	26.26%	0.44
	N of English words	Higher	20.58%	0.41
		Lower	31.82%	0.47
	N of Label naming	Higher	30.03%	0.46
		Lower	26.83%	0.44
	Phrase	YES	35.09%	0.48
		NO	27.02%	0.44
Prosodic cues	Speaking rate	Higher	26.13%	0.44
		Lower	29.01%	0.45
	Pitch	Higher	26.43%	0.44
		Lower	28.76%	0.45
	Pauses	Higher	25.54%	0.44
		Lower	29.31%	0.46
Nonverbal cues	Gestures	YES	27.78%	0.45
		NO	27.52%	0.44
	Iconic rate	Higher	30.03%	0.46
		Lower	26.80%	0.44
	Point rate	Higher	22.70%	0.42
		Lower	29.20%	0.45
	Beat rate	Higher	28.70%	0.45
		Lower	27.50%	0.45
	Interactive rate	Higher	31.13%	0.46
		Lower	27.07%	0.44
	Gaze	YES	28.24%	0.45
		NO	24.52%	0.43

Note: For descriptive purposes, variables are categorized into 'Higher ' and 'Lower' groups based on their mean, with scores above and below the mean, respectively.

First, as shown in Figure 1, students who had a better working memory (β = 0.17, p = .038), higher English proficiency (β = 0.42, p < .001), a higher degree of liking the teacher (β = 0.20, p = .041) positively predicted their learning outcomes. After controlling for these significant students' internal factors, teachers' wordings, prosody, and gestures still had significant effects. For the linguistic features, teachers asking more questions (β = 0.33, p = .02), using fewer English words (β = -0.56, p = .050), teaching phrases (β = 1.05, p = .048), and speaking shorter in total during teaching a vocabulary in a clip (β = -0.03, p = .09, two-tailed) enhanced student learning outcomes. Prosodically, teachers

who had a slower speaking rate (β = -0.80, p = .002) and a lower pausing rate (β = -3.67, p = .02) predicted students' better vocabulary learning outcomes. As for gestures, a word in a clip where a teacher gestured at least once was better learned than that of a teacher who made no gesture (β = 0.86, p = .015). Further analysis of the frequency of different types of gestures showed that it was mainly the iconic gesture rate (β = 0.21, p = .027) that facilitated students' unknown word learning (Figure 2).

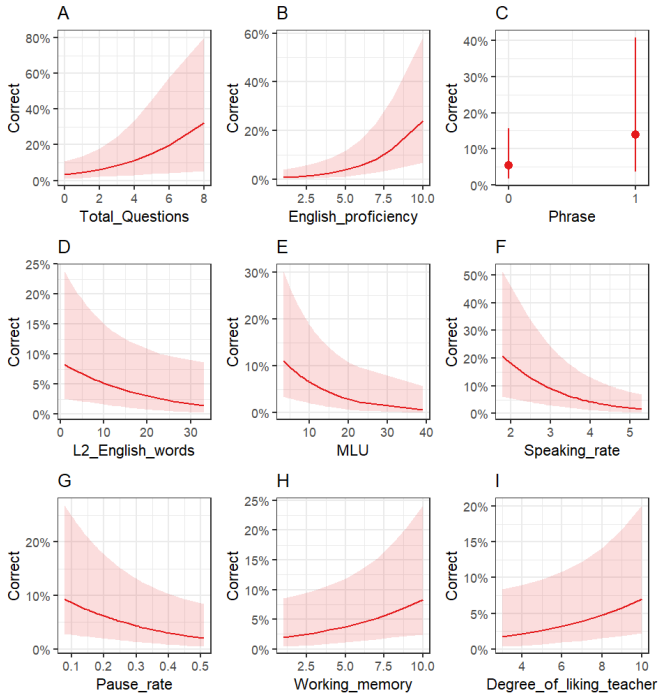

Figure 1. The predicted effects of different factors (A-I) on students' unknown vocabulary learning outcomes, with 95% CI (shaded bands).

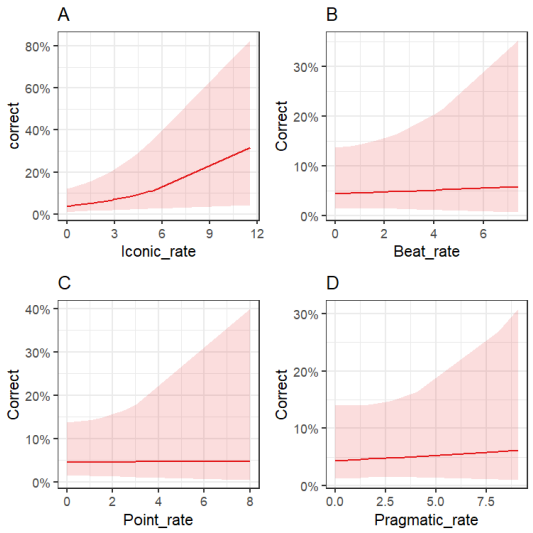

Figure 2. Predicted effects of different types of gesture rate per 100 words (A-D), with 95% CI (shaded bands).

Discussion

This is the first comprehensive study that simultaneously investigates the collective impact of teachers' linguistic features, speech prosody and gestures on student learning outcomes in naturalistic L2 teaching while controlling for various factors. We found that multimodal cues significantly enhance students' L2 vocabulary learning, with implications for educators, linguists and cognitive scientists.

Recent research emphasized the importance of verbal and multimodal behaviors (e.g., higher pitch, faster speaking rate, points) in adult conceptual learning (Edwards et al., 2024). Additionally, multimodal cues (pitch, indexical, and iconic gestures) predicted English adults' ability to learn the names of unknown objects (Cabiddu et al., 2024). Unlike these two studies focusing on native speakers, our research is the first to comprehensively quantify the role of various multimodal cues in L2 vocabulary learning while accounting for significant individual differences.

Linguistic features

First, the results suggest better learning outcomes correlate with shorter MLU in vocabulary instruction. This is consistent with some findings in the developmental research where there was a negative correlation between parental MLU and vocabulary size (Dong et al., 2021). Interestingly, while statistical learning research discusses the benefits of additional input in strengthening perceived patterns (Aslin et al., 1998), speakers use shorter utterances to children than to adults (Zhang & Gu, 2023). Similarly, in our study, we found that shorter utterances have a positive impact on adult L2 vocabulary learning, implying that shorter utterances may better align with the learner's current linguistic and cognitive abilities. This finding resonates with Reigeluth's Elaboration Theory (1992), which advocates for instruction that emphasizes minimizing cognitive load by focusing on essential cues to improve learning efficiency.

Second, students had better learning outcomes when the vocabulary instruction had fewer L2 English words and emphasised "phrase" learning. This supports the importance of familiarity in vocabulary acquisition (Qiu & Lo, 2017). While previous studies suggested that learners speaking more L2 aids students' phonetics (Trofimovich et al., 2007, 2012) and listening comprehension (Lee & Levine, 2020). Our findings show significant advantages of using L1 for education outcomes, aligning with findings on the efficacy of L1 in facilitating clear understanding through L1 translation equivalents (Leeser, 2007; Lee & Macaro, 2013). Our study indicates that explanations in Chinese coupled with "phrase" study notably improve students' processing and retention of L2 vocabulary, likely due to decreased cognitive load and connections to familiar concepts. Therefore, we underscore the value of using L1 and incorporating "phrases" in vocabulary instruction.

Third, teachers' use of questions has benefits for student learning outcomes. Research indicates that questioning significantly enhances student learning outcomes in a science study (McCarthy et al., 2016; Tofade, 2013). Asking

questions may draw students' attention and encourage deeper thinking, helping them clarify their thoughts and consider diverse viewpoints (Van & Minstrell, 1997). This is especially true for open-ended questions (e.g., "What is a compliment?"), which stimulates greater cognitive efforts and encourages deeper content processing (Wasik & Bond, 2001). Furthermore, asking questions promotes active thinking and enhances students' engagement with the material. Although students may not always appreciate questioning (Zhou & Gu, 2024), this method is key for capturing their attention and developing critical and creative thinking skills (Nappi, 2017).

Aslin and Newport (2012) show that statistical learning mechanisms can detect regularities in patterns, such as repetitions in an AAB sequence. However, our study found that the frequency of repetitions did not significantly affect learning outcomes. Thus, merely repeating a word may not substantially aid in understanding its semantic content. This study proposes that effective vocabulary acquisition involves more than just exposure to the word form but requires engaging with the word in meaningful contexts. For instance, Webb (2007) argues that encountering a new word in varied contexts can enhance the understating of its use and nuances, suggesting that the quality of exposure is more critical than quantity.

Prosodic cues

Although pitch had no observed effect, we found that a slower speaking rate positively influenced students' learning. Previous studies have mainly focused on the impact of speaking rate on word segmentation or vocabulary size in child-directed speech (Han et al., 2024) and phoneme recognition (Krause & Braida, 2004), but they often did not extend to the realm of word learning directly, with research like Cabiddu et al. (2024) and Shi et al. (2023) indicating that the effect of speaking rate on vocabulary acquisition remained unclear. Our research demonstrates the direct benefit of speaking rate on vocabulary learning, specifically in tertiary L2 learners. This unique contribution highlights that a moderated speaking pace can significantly enhance the acquisition of new vocabulary, even in adults. Francis and Nusbaum (1996) observed that a slower speech rate reduces cognitive load in L1 speaking, and this approach not only enhances immediate comprehension but also supports long-term retention of the learned material, as confirmed by Nanjo and Kawahara (2002). Our findings extend these insights, demonstrating that a slower speech rate also benefits learning in L2 contexts.

Furthermore, we found that students learned better when teachers had a lower rate of pauses, indicating that a slower speaking rate does not mean having disfluencies in speech. While earlier studies suggested that lecture fluency mainly affects students' perceptions rather than their learning (Carpenter et al., 2013, 2016), recent findings indicate that fluency positively affects learning performance (Wilford et al., 2020). Our study aligns with this emerging perspective, showing that fluency enhances vocabulary comprehension, particularly a lower rate of speech pauses. Contrary to the belief that slower speech compromises fluency, we found that a fluent delivery enhances learning even if slower. This acknowledges fluency's role in effective teaching and redefines the relationship between speech rate and fluency in learning contexts.

Gestures

Our research highlights the effectiveness of L2 learning, demonstrating that students better recall words from a video clip if the teacher made at least one gesture than without gestures (Huang et al., 2019; Oppici et al., 2023). This finding aligns with the Dual Coding Theory (Paivio, 2013), which posits that learning is more efficient when it engages verbal and visual inputs. In the classroom, teachers' gestures can also establish a common understanding among students for a clear grasp of the instructional content (Alibali et al., 2019).

Despite the roles of gestures in EFL teaching is not new (Lin, 2021; McCafferty, 2002, 2006), we show for the first time quantitatively the effect of gestures on L2 vocabulary learning in naturalistic teaching, controlling for many important factors. Our study emphasizes the unique role of iconic gestures (rather than other gesture types) in vocabulary learning. Previous studies have underscored the effectiveness of iconic gestures in connecting new words with established meanings (Kelly et al., 2009) and enhancing long-term retention (Aussems & Kita, 2019; Khanukaeva, 2014). Iconic gestures can facilitate understanding through mechanisms of "displacement," "referentiality," and "embodiment" (Perniss & Vigliocco, 2014). Our findings extend these insights to adult L2 learners, showing that iconic gestures significantly improve vocabulary acquisition, even after accounting for many key individual differences.

Conclusion

This study quantitatively analyzed how teachers' multimodal cues, alongside students' individual characteristics, jointly impact L2 word learning. Our findings highlight that a combination of shorter utterances, more questions, the use of gestures, and careful modulation of prosody can significantly enhance student learning outcomes. These results contribute to a better understanding of multimodal language processing and learning and suggest that adapting teaching strategies to include varied multimodal cues can optimize student engagement and comprehension. Future studies might further investigate the broader applicability of our results in enhancing language teaching methodologies and learning experiences.

Acknowledgements

We thank the teachers and students who participated in this study. The work was supported by The Humanity and Social Science Research Foundation from the Ministry of Education of China (19YJC740001).

References

Alibali, M. W., Nathan, M. J., Boncoddo, R., & Pier, E. (2019). Managing common ground in the classroom: teachers use gestures to support students' contributions to classroom discourse. *ZDM*, 51, 347-360.

Arealme.(n.d.) *Memory Test*. Retrieved from https://www.arealme.com/memory-test/cn/

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4), 321-324.

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, 21(3), 170-176.

Aussems, S., & Kita, S. (2019). Seeing iconic gestures while encoding events facilitates children's memory of these events. *Child Development*, *90*(4), 1123-1137.

Baafi, R. K. A. (2020). Teacher-student relationship and student learning outcomes in senior public secondary schools in Ghana. *European Journal of Education Studies*. 6(12), 147-161.

Baese-Berk, M. M., Heffner, C. C., Dilley, L. C., Pitt, M. A., Morrill, T. H., & McAuley, J. D. (2014). Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychological Science*, *25*(8), 1546-1553.

Bavelas, J. B., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. *Personality and Social Psychology Bulletin*, *21*(4), 394-405.

Berthier, M. L., & Lambon Ralph, M. A. (2014). Dissecting the function of networks underpinning language repetition. *Frontiers in Human Neuroscience*, *8*, 727.

Boersma, Paul & Weenink, David (2023). *Praat: doing phonetics by computer* [Computer program]. Version 6.3.08, retrieved 10 February 2023 from http://www.praat.org/.

Bowers, E. P., & Vasilyeva, M. (2011). The relation between teacher input and lexical growth of preschoolers. *Applied Psycholinguistics*, *32*(1), 221-241.

Cabiddu, F., Edwards, C., Hill-Payne, H., Donnellan, E., Gu, Y. & Vigliocco, G. (2024). What predicts adult word learning in naturalistic interactions? A corpus study. *Proceedings of the Annual Meeting of the Cognitive Science Society.*

Candry, S., Deconinck, J., & Eyckmans, J.(2018). Written repetition vs. oral repetition: Which is more conducive to L2 vocabulary learning?. *Journal of the European Second Language Association*, *2*(1) 72-82.

Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review*, *20*, 1350-1356.

Carpenter, S. K., Mickes, L., Rahman, S., & Fernandez, C. (2016). The effect of instructor fluency on students' perceptions of instructors, confidence in learning, and actual learning. *Journal of Experimental Psychology: Applied*, *22*(2), 161.

Clark, J., & Trofimovich, P. (2016). L2 vocabulary teaching with student-and teacher-generated gestures: A classroom perspective. *TESL Canada Journal*, *34*(1), 1-24.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*(2), 1047-1058.

Cowan, N. (2012). *Working memory capacity*. New York, NY: Psychology Press

Dong, S., Gu, Y., & Vigliocco, G. (2021). The impact of child-directed language on children's lexical development. In *Proceedings of the Annual Meeting of the Cognitive Science Society, 43*.

Donnellan, E., Jordan-Barros, A., Theofilogiannakou, N., Brekelmans, G., Murgiano, M., Motamedi, Y., ... & Vigliocco, G. (2023). The impact of caregivers' multimodal behaviours on children's word learning: A corpus-based investigation. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*.

Drijvers, L., & Holler, J. (2023). The multimodal facilitation effect in human communication. *Psychonomic Bulletin & Review*, 30(2), 792-801.

Duong, S., Bachman, H. J., Votruba-Drzal, E., & Libertus, M. E. (2021). What's in a question? Parents' questions use in dyadic interactions and the relation to preschool-aged children's math abilities. *Journal of Experimental Child Psychology,* 211, 105213.

Edwards, C., Cabiddu, F., Hill-Payne, H., D'Estalenx, Q., Donnellan, E., Gu, Y. & Vigliocco, G. (2024). The impact of speakers' multimodal behaviours on adults' learning of semantic information: A corpus-based investigation. *Proceedings of the Annual Meeting of the Cognitive Science Society.*

Edstrom, A. (2009). Teacher reflection as a strategy for evaluating L1/L2 use in the classroom. *Babylonia*, *1*(09), 12- 15.

Filippi, P., Gingras, B., & Fitch, W. T. (2014). Pitch enhancement facilitates word learning across visual contexts. *Frontiers in Psychology*, 5, 1468.

Francis, A. L., & Nusbaum, H. C. (1996, October). Paying attention to speaking rate. In *Proceeding of Fourth International Conference on Spoken Language Processing*. ICSLP'96 (Vol. 3, pp. 1537- 1540). IEEE.

Fredrick, W. C., & Walberg, H. J. (1980). Learning as a function of time. *The Journal of Educational Research*, 73(4), 183-194.

Han, M. (2019). *The role of prosodic input in word learning: A cross-linguistic investigation of Dutch and Mandarin Chinese infant-directed speech* (Doctoral dissertation, LOT).

Han, M., DE JONG NH, R. KAGER, R. (2024). Relating the prosody of infant-directed speech to children's vocabulary size. *Journal of Child Language*, 51(1), 217-233.

Hart, B., & Risley, T. R. (1996). Meaningful differences in the everyday experience of young American children. *Community Alternatives*, 8, 92-93.

Huang, X., Kim, N., & Christianson, K. (2019). Gesture and vocabulary learning in a second language. *Language Learning*, 69(1), 177- 197.

Hudson, N. (2011). *Teacher gesture in a post-secondary English as a second language classroom: A sociocultural approach* (Doctoral dissertation, University of Nevada, Las Vegas).

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. *Cognition and Second Language Instruction*, 258- 286.

Jiang, X. (2011). The role of first language literacy and second language proficiency in second language reading comprehension. *The Reading Matrix*, 11(2), 177-190.

Kelly, S. D., & Lee, A. L. (2012). When actions speak too much louder than words: Hand gestures disrupt word learning when phonetic demands are high. *Language and Cognitive Processes*, 27(6), 793-807.

Kelly, S. D., McDevitt, T., & Esch, M. (2020). Brief training with co-speech gesture lends a hand to word learning in a foreign language. In *Speech Accompanying-Gesture* (pp. 313-334). Psychology Press.

Khanukaeva, A. (2014). *The effects of iconic gestures on L2 vocabulary learning in a Norwegian primary school* (Master' s thesis, University of Stavanger, Norway).

Krause, J. C., & Braida, L.D.(2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, 115(1), 362-378.

Krishnan, S., Alcock, K. J., Carey, D., Bergström, L., Karmiloff- Smith, A., & Dick, F. (2017). Fractionating nonword repetition: The contributions of short-term memory and oromotor praxis are different. *PLoS One*, 12(7), e0178356.

Kuang, Z., Wang, F., Xie, H., Mayer, R. E., & Hu, X. (2023). Effect of the instructor's eye gaze on student learning from video lectures: Evidence from two three-level meta-analyses. *Educational Psychology Review*, 35(4), 109.

Kushch, O., & Prieto Vives, P. (2016). The effects of pitch accentuation and beat gestures on information recall in contrastive discourse. Barnes J, Brugos A, Shattuck-Hufnagel S, Veilleux N, editors. *Speech Prosody* 2016; Boston, United States of America. p. 922-5. DOI: 10.21437/SpeechProsody. 2016-189.

Kushch, O., Igualada, A., & Prieto, P. (2018). Prominence in speech and gesture favour second language novel word learning. *Language, Cognition and Neuroscience*, 33(8), 992- 1004.

Lajevardi, N., Narang, N. S., Marcus, N., & Ayres, P. (2017). Can mimicking gestures facilitate learning from instructional animations and static graphics?. *Computers & Education*, 110, 64-76.

Larsen-Freeman, D., & Anderson, M. (2013). *Techniques and principles in language teaching 3rd edition-Oxford handbooks for language teachers*. Oxford University Press.

Lee, O. S. (2023). Implicit Statistical Learning in L2 Sentence Processing: Individual Cognitive Differences *Journal of Psycholinguistic Research*, 52(4), 1037-1060.

Lee, J. H., & Macaro, E. (2013). Investigating age in the use of L1 or English-only instruction: Vocabulary acquisition by Korean EFL learners. *The Modern Language Journal*, 97(4), 887-901.

Lee, J. H., & Levine, G. S. (2020). The effects of instructor language choice on second language vocabulary learning and listening comprehension. *Language Teaching Research*, 24(2), 250-272.

Leeser, M. J. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning*, 57(2), 229- 270.

Lin, Y. L. (2021). Gestures as scaffolding for L2 narrative recall: The role of gesture type, task complexity, and working memory. *Language Teaching Research*, 1-23.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, 21, 861-883.

Malécot, A., Johnston, R., & Kizziar, P.A. (1972). Syllabic rate and utterance length in French. *Phonetica*, 26(4), 235–251.

Martirosyan, N. M., Saxon, D. P., & Wanjohi, R. (2014). Student satisfaction and academic performance in Armenian higher education. *American International Journal of Contemporary Research*, 4(2), 1-5.

McCafferty, S. G. (2002). Gesture and creating zones of proximal development for second language learning. *The Modern Language Journal*, 86(2), 192-203.

McCafferty, S.G. (2006). Gesture and the materialization of second language prosody. *International Review of Applied Linguistics in Language Teaching*, 44(4), 195-209.

McCarthy, P., Sithole, A., McCarthy, P., Cho, J. P., & Gyan, E. (2016). Teacher questioning strategies in mathematical classroom discourse: A case study of two grade eight teachers in Tennessee, USA. *Journal of Education and Practice*, 7(21), 80-89.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

Mirzaei, A., Azizi Farsani, M., & Chang, H. (2023). Statistical learning of L2 lexical bundles through unimodal, bimodal, and multimodal stimuli. *Language Teaching Research*, 1-25.

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology,* 18(1), 118-141.

Moser, J., Batterink, L., Hegner, Y. L., Schleger, F., Braun, C., Paller, K. A., & Preissl, H. (2021). Dynamics of nonlinguistic statistical learning: From neural entrainment to the emergence of explicit knowledge. *NeuroImage*, 240, 118378.

Munro, M. J., & Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning*, 48(2), 159-182.

Nanjo, H., & Kawahara, T. (2002, May). Speaking-rate dependent decoding and adaptation for spontaneous lecture speech recognition. In 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 1, pp. I-725). *IEEE*.

Nappi, J. S. (2017). The importance of questioning in developing critical thinking skills. *Delta Kappa Gamma Bulletin*, 84(1), 30.

Oppici, L., Mathias, B., Narciss, S., & Proske, A. (2023). Benefits of enacting and observing gestures on foreign language vocabulary learning: A systematic review and meta-analysis. *Behavioral Sciences*, 13(11), 920.

Osorio, S., Straube, B., Meyer, L., & He, Y. (2024). The role of co-speech gestures in retrieval and prediction during naturalistic multimodal narrative processing. *Language, Cognition and Neuroscience*, 39(3), 367-382.

Paivio, A. (2013). *Imagery and verbal processes.* New York, NY: Psychology Press

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130300.

Pujadas, G., & Muñoz, C. (2020). Examining adolescent EFL learners' TV viewing comprehension through captions and subtitles. *Studies in Second Language Acquisition*, 42(3), 551-575.

Qiu, X., & Lo, Y. Y. (2017). Content familiarity, task repetition and Chinese EFL learners' engagement in second language use. *Language Teaching Research*, 21(6), 681-698.

R Core Team (2020). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, https://www.R-project.org.

Rauf, A. (2018). Students' attitude towards teachers' use of code-switching and its impact on learning English. *International Journal of English Linguistics*, 8(1), 212-218.

Raneri, D., Von Holzen, K., Newman, R., & Ratner, N. B. (2020). Change in maternal speech rate to preverbal infants over the first two years of life. *Journal of Child Language*, 47(6), 1263-1275.

Reigeluth, C. M. (1992). Commentary: Elaborating the elaboration theory. *Educational Technology Research and Development*, 80-86.

Rohrer, P. L., Delais-Roussarie, E., & Prieto, P. (2020). Beatgestures for comprehension and recall: Differential effects of language learners and native listeners. *Frontiers in Psychology,* 11, 575929.

Samad, A. A., Etemadzadeh, A., & Far, H. R. (2012). Motivation and language proficiency: Instrumental and integrative aspects. *Procedia-Social and Behavioral Sciences*, 66, 432-440.

Saito, K., Dewaele, J. M., Abe, M., & In'nami, Y. (2018). Motivation, emotion, learning experience, and second language comprehensibility development in classroom settings: A cross-sectional and longitudinal study. *Language Learning*, 68(3), 709- 743.

Sharma, K., Alavi, H. S., Jermann, P., & Dillenbourg, P. (2016, April). A gaze-based learning analytics model: in-video visual feedback to improve learner's attention in MOOCs. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 417-421).

Shi, J., Gu, Y., & Vigliocco, G. (2023). Prosodic modulations in child-directed language and their impact on word learning. *Developmental Science*, 26(4), e13357.

Sikveland, R.O., Solem, M. S., & Skovholt, K. (2021). How teachers use prosody to guide students towards an adequate answer. *Linguistics and Education*, 61, 100886.

Smith,S.A.,& Murphy,V.A.(2015).Measuring productive elements of multi-word phrase vocabulary knowledge among children with English as an additional or only language. *Reading and Writing*, 28, 347-369.

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustical Society of America*, 128(1), 389-400.

Stam, G., Tellier, M., & Bigi, B. (2012). *Handling language: The gestures of future foreign language teachers*. Faculty Publications.

Stam, G., & Tellier, M. (2022). Gesture helps second and foreign language learning and teaching. *Gesture in language: Development across the lifespan*, 336–363.

Sun, H., & Verspoor, M. (2022). Mandarin vocabulary growth, teacher qualifications and teacher talk in child heritage language learners. *International Journal of Bilingual Education and Bilingualism*, *25*(6), 1976-1991.

Takahashi, C., Kao, S., Baek, H., Yeung, A. H., Hwang, J., & Broselow, E. (2018). Native and non-native speaker processing and production of contrastive focus prosody. *Proceedings of the Linguistic Society of America*, *3*(1), 35-1.

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, *8*(2), 219-235.

Tofade, T., Elsner, J., & Haines, S. T. (2013). Best practice strategies for effective use of questions as a teaching tool. *American Journal of Pharmaceutical Education*, *77*(7), 155.

Tompkins, V., Bengochea, A., Nicol, S., & Justice, L. M. (2017). Maternal inferential input and children's language skills. *Reading Research Quarterly*, *52*(4), 397-416.

Trofimovich, P., Gatbonton, E., & Segalowitz, N. (2007). A dynamic look at L2 phonological learning: Seeking processing explanations for implicational phenomena. *Studies in Second Language Acquisition*, *29*(3), 407-448.

Trofimovich, P., Collins, L., Cardoso, W., White, J., & Horst, M. (2012). A frequency-based approach to L2 phonological learning: Teacher input and student output in an intensive ESL context. *Tesol Quarterly*, *46*(1), 176-187.

Van Zee, E., & Minstrell, J. (1997). Using questioning to guide student thinking. *The Journal of the Learning Sciences*, 6(2), 227-269.

Wang, J., Gao, Y., & Cui, Y. (2023). Classroom gesture instruction on second language learners' academic presentations: Evidence from Chinese intermediate English learners. *Journal of English for Academic Purposes*, *66*, 101304.

Wasik, B. A., & Bond, M. A. (2001). Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology*, *93*(2), 243.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied linguistics*, *28*(1), 46-65.

Wilford, M. M., Kurpad,N.,Platt, M.,& Weinstein-Jones, Y. (2020). Lecturer fluency can impact students' judgments of learning and actual learning performance. *Applied Cognitive Psychology*, *34*(6), 1444-1456.

Zhang, Y., Ding, R., Frassinelli, D., Tuomainen, J., Klavinskis-Whiting, S., & Vigliocco, G. (2023). The role of multimodal cues in second language comprehension. *Scientific Reports*, *13*(1), 20824.

Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I., & Vigliocco, G. (2021). More than words: Word predictability, prosody, gesture and mouth movements in natural language comprehension. *Proceedings of the Royal Society B*, 288(1955), 20210500.

Zhang, Y., & Gu, Y. (2023). A recipient design in multimodal language on TV: A comparison of child-directed and adult-directed broadcasting. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).

Zhou, J., & Gu, Y (2024). Unraveling students' liking of teachers: The impact of multimodal cues during L2 English vocabulary teaching. *Proceedings of Speech Prosody 2024*. Netherlands: Leiden.