

COMMENTARY



Connecting multimodal learning processes and outcomes in CALL: Introduction to the special issue

Bronson Hui, University of Maryland, College Park

Matt Kessler, University of South Florida

APA Citation: Hui, B., & Kessler, M. (2025). Connecting multimodal learning processes and outcomes in CALL: Introduction to the special issue. *Language Learning & Technology*, 29(3), 1–5.

<https://doi.org/10.64152/10125/73644>

Introduction

In recent years, there has been a spike in interest in *multimodality*, a concept which refers to the use of two or more *modes* (i.e., meaning-making resources including those that are linguistic, aural, visual, gestural, or spatial in nature) for communicative purposes (Jewitt, 2006; Kress, 2010). Multimodality itself has long existed among human societies, and it has even been an object of inquiry among many academic fields for decades (for example, see Halliday, 1978; Jewitt et al., 2016). More recently, researchers in the field of Second Language Acquisition (SLA) have begun to examine the extent to which multimodality plays a role in fostering processes related to second language (L2) learning and teaching. In particular, a sizable amount of scholarship within SLA has focused on two key aspects, involving either (a) writing or (b) receiving multimodal input (e.g., captioned media and reading-while-listening). In terms of the former, writing-focused studies involving multimodality have predominantly been situated in classroom contexts, with researchers examining the practical affordances and constraints of adopting multimodal composing tasks in their pedagogies (e.g., Dzekoe, 2017; Hafner, 2015; Jiang, 2017; Kim et al., 2022; Shin et al., 2020; also see Lim & Kessler, 2024 for a research timeline). Conversely, in terms of receiving multimodal input, inquiries have often focused on the cognitive processing of multimodal input and its effects on comprehension, vocabulary acquisition, and in some instances, grammar learning (e.g., Montero Perez, 2022; Muñoz et al., 2023; Pellicer-Sánchez et al., 2020; Puimège et al., 2023; The TwiLex Group, 2024; Vanderplank, 2016).

As interest in multimodality has expanded, so, too, have advancements in technology. As a result, an increasing number of computer-assisted language learning (CALL) scholars have continued existing lines of inquiry and explored new developments related to the digital, computer-mediated nature of multimodality (e.g., Gass et al., 2019; Kessler, 2024; Li, 2021; Li & Jiang, 2024; Sydorenko et al., 2024; Zhang & Yu, 2024). That is, more and more researchers are leveraging emerging internet-based programs, software, mobile apps, and digital tools as a means of understanding the intersections of multimodality and SLA. Despite this growing interest in the field at large, there is a paucity of research within CALL, particularly when it comes to investigating the connections between L2 learning processes and outcomes. Most of the literature to date has tended to independently examine issues of multimodal processes *or* the effects of different modes, respectively. Thus, CALL research is needed that adopts established theories to examine how L2 learners engage with different multimodal tasks/activities, *and* how learners' engagement leads to (or does not lead to) a myriad of learning outcomes (e.g., vocabulary gains, writing gains, increased metacognition, and more).

Overview of the Special Issue

In response to such needs, the current special issue on *Connecting Multimodal Learning Processes and Outcomes in CALL* was developed. Specifically, the studies featured in this issue interrogate links between L2 learning processes and outcomes by examining diverse topics of inquiry, and also through adopting a variety of research methods. For this special issue, we have organized the articles into three major themes/areas. These themes include studies involving (a) *vocabulary*, (b) *digital multimodal composing*, and (c) *speech and pragmatics*. Contributions within these respective areas are briefly described in what follows.

The first theme in this special issue involves studies which focus on links between multimodality and *vocabulary* learning. There are five different studies that correspond to this theme. In the first study by Tuzcu (2025), the author adopted eye-tracking technology combined with verbal reports. These methods were used to investigate how additional auditory information (i.e., the aural rendition of text) influences attention to and awareness of target words in the context of vocabulary learning during reading-while listening, which in turn has an impact on vocabulary gains. Also within the domain of vocabulary through reading-while-listening, Malone (2025) explores how learners interact with input under reading-only and reading-while-listening conditions. Using eye-movement data, Malone captured visit counts and reading times per visit, which he then associated with three offline vocabulary measures. The third study by Sun et al. (2025) also explores reading-vocabulary links, but in the context of children's reading of eBooks. In the study, Sun et al. examine the (lack of) impact of nonverbal sound effects on directing the reader's attention to target words. The authors discuss how the provision of multimodal input may not necessarily be beneficial, as it is conditional on factors such as the relevance of the information and the proficiency of the learner. The fourth paper involving vocabulary is concerned with the role of on-screen texts in the viewing of videos. In this study by Wang (2025), the author tests the extent to which processing behavior (e.g., reading ahead of the audio or skipping an unfamiliar word) is a function of on-screen text types: L2 captions, L1 subtitles, and bilingual subtitles. The author highlights the complexity of the multimodal processing task where the learner juggles between imagery, audio, and written text in two languages. Finally, in Wu et al. (2025), the authors examine vocabulary learning from different types of glossing (i.e., L2 textual gloss, pictorial gloss, and pictorial plus L2 textual gloss). The authors were also interested in how look-up behavior and individual differences factors (i.e., working memory capacity) might shape vocabulary learning in repeated exposures. Notably, for all the studies in this section of the special issue, eye tracking is a key tool used to capture attention in the learning process; however, other methods, such as verbal reports and system logs revealing look-up behavior, offer insights into the learning journey that learners go through.

The second theme in this special issue involves studies on the topic of *digital multimodal composing*. In this theme, there are two studies that explore different aspects of digital multimodal composing (DMC). In the first study by Xu et al. (2025), the authors investigate the potential of a DMC task to enhance aspects of student writers' metacognition. Their mixed methods study leveraged both an experimental group and a comparison group who completed different writing tasks (i.e., a multimodal photo essay vs. a monomodal narrative). The researchers then collected different data to examine aspects of learners' metacognitive knowledge before, during, and after the intervention to understand the relationship between learning processes and enhanced metacognition. In the second study on DMC, Stell (2025) also adopts mixed methods; however, her study employs Activity Theory as a lens for exploring the writing processes and outcomes of learners who worked collaboratively on a source-based multimodal writing activity. Through data such as keystroke logging, audio recordings of pair talk, and students' written products, Stell details how instances of pair talk (via language-related episodes) and digital tools resulted in different learning opportunities and mediated participants' writing strategies.

The third and final theme within this special issue covers studies on *speech and pragmatics*. There are two studies in this section. In the first study by Montero Perez and Pattemore (2025), the authors investigate the extent to which captions (visualizing the auditory input in viewing) can improve learners'

capacity to segment incoming speech streams. Using a shadowing task where participants repeated back sentences, the authors test an important psycholinguistic account for the widely reported benefits of captioned media for comprehension and vocabulary learning – namely, that the texts on the screen assist phonological decoding, hence allowing for processing at the higher levels. In the second study, Peng and Lei (2025) explore interactions in different modes of synchronous computer-mediated communication (video chat vs. text chat). Specifically, they examine if different modes might afford opportunities to explicitly discuss aspects of pragmatics (learning processes), and eventually influence the frequency of use of certain pragmatic strategies (learning outcomes).

Conclusion

The topic of multimodality is becoming increasingly relevant in applied linguistics and education, particularly as a sizable amount of our interactions and daily activities continue to become digitized. In turn, understanding the connections between multimodality, L2 learning processes, and L2 learning outcomes will also become increasingly critical. Of course, the contributions to this special issue represent only a small fraction of the potential avenues of study in this domain. However, we ultimately hope that the articles in this issue will help spur discussions on this topic, both within the domain of CALL and beyond. As a final remark, many of the contributors in this special issue have generously shared their materials, data, and analyses in the spirit of open science and scholarship (for more, see the links embedded within individual articles). We appreciate their efforts in helping the field gain the much-needed momentum towards more credible and reproducible research.

References

- Dzekoe, R. (2017). Computer-based multimodal composing activities, self-revision, and L2 acquisition through writing. *Language Learning & Technology*, 21(2), 73–95. <https://dx.doi.org/10125/44612>
- Gass, S., Winke, P., Isbell, D. R., & Ahn, J. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning & Technology*, 23(2), 84–104. <https://doi.org/10125/44684>
- Hafner, C. (2015). Remix culture and English language teaching: The expression of learner voice in digital multimodal composition. *TESOL Quarterly*, 49(3), 486–509. <https://doi.org/10.1002/tesq.238>
- Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. University Park Press.
- Jewitt, C. (2006). *Technology, literacy and learning A multimodal approach*. Routledge.
- Jewitt, C., Bezemer, J., & O'Halloran, K. (2016). *Introducing multimodality*. Routledge.
- Jiang, L. (2017). The affordances of digital multimodal composing for EFL learning. *ELT Journal*, 71(4), 413–422. <https://doi.org/10.1093/elt/ccw098>
- Kessler, M. (2024). *Digital multimodal composing: Connecting theory, research and practice in second language acquisition*. Multilingual Matters.
- Kim, Y., Kang, S., Nam, Y., & Skalicky, S. (2022). Peer interaction, writing proficiency, and the quality of collaborative digital multimodal composing task: Comparing guided and unguided planning. *System*, 106, Article 102722. <https://doi.org/10.1016/j.system.2022.102722>
- Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.
- Li, M. (2021). *Researching and teaching second language writing in the digital age*. Palgrave Macmillan.

- Li, M., & Jiang, L. (2024). Digital multimodal pedagogy for teacher training. In R. Hampel & U. Stickler (Eds.), *The Bloomsbury handbook of language learning and technology* (pp. 446–459). Bloomsbury Publishing.
- Lim, J., & Kessler, M. (2024). Multimodal composing and second language acquisition. *Language Teaching*, 57(2), 183–202. <https://doi.org/10.1017/S0261444823000125>
- Malone, J. (2025). Multimodal effects on processing and learning contextualized L2 vocabulary. *Language Learning & Technology*, 29(3), 49–69. <https://doi.org/10.64152/10125/73647>
- Montero Perez, M. (2022). Second or foreign language learning through watching audio-visual input and the role of on-screen text. *Language Teaching*, 55(2), 163–192. <https://doi.org/10.1017/S0261444821000501>
- Montero Perez, M., & Pattenmore, 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>
- Muñoz, C., Pujadas, G., & Pattenmore, A. (2023). Audio-visual input for learning L2 vocabulary and grammatical constructions. *Second Language Research*, 39(1), 13–37. <https://doi.org/10.1177/02676583211015797>
- Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Serrano, R., & Llanes, Á. (2020). Young learners' processing of multimodal input and its impact on reading comprehension: an eye-tracking study. *Studies in Second Language Acquisition*, 42(3), 577–598. <https://doi.org/10.1017/S0272263120000091>
- Peng, Y., & Lei, Y. (2025). Investigating pragmatic learning opportunities and outcomes in different SCMC modes. *Language Learning & Technology*, 29(3), 226–246. <https://doi.org/10.64152/10125/73654>
- Puimège, E., Montero Perez, M., & Peters, E. (2021). Promoting L2 acquisition of multiword units through textually enhanced audiovisual input: An eye-tracking study. *Second Language Research*, 39(2), 471–492. <https://doi.org/10.1177/02676583211049741>
- Shin, D., Cimasko, T., & Yi, Y. (2020). Development of metalanguage for multimodal composing: A case study of an L2 writer's design of multimedia texts. *Journal of Second Language Writing*, 47, Article 100714. <https://doi.org/10.1016/j.jslw.2020.100714>
- Stell, A. (2025). Exploring multimodality in technology-mediated collaborative writing: A lens through Activity Theory. *Language Learning & Technology*, 29(3), 182–204. <https://doi.org/10.64152/10125/73652>
- Sun, H., Roberts, A., Tan, J., Loh, J., & Moh, Y. C. Y. (2025). Language proficiency over sound effects in children's eBook vocabulary learning. *Language Learning & Technology*, 29(3), 135–156. <https://doi.org/10.64152/10125/73650>
- Sydorenko, T., Cárdenas-Claros, M. S., Huntley, E., & Montero Perez, M. (2024). Audiovisual input in language learning: Teachers' perspectives. *Language Learning & Technology*, 28(1), 1–25. <https://hdl.handle.net/10125/73587>
- The TwiLex Group. (2024). First language effects on incidental vocabulary learning through bimodal input: A multisite, preregistered, and close replication of Malone. *Studies in Second Language Acquisition*, 46(5), 1413–1438. <https://doi.org/10.1017/S0272263124000275>
- Tuzcu, A. (2025). The effects of input modality on novel L2 vocabulary processing during reading: A closer look into attention and awareness. *Language Learning & Technology*, 29(3), 30–48. <https://doi.org/10.64152/10125/73646>

- Vanderplank, R. (2016). *Captioned media in foreign language learning: Subtitles for the deaf and hard-of-hearing as tools for language learning*. Palgrave Macmillan.
- Wang, A. (2025). The integration of auditory and textual input in incidental vocabulary learning from subtitled viewing: An eye-tracking study. *Language Learning & Technology*, 29(3), 70–91.
<https://doi.org/10.64152/10125/73648>
- Wu, Y., Liu, X., Zhao, Y., Chen, X., & Gao, J. (2025). Multimodal glosses revisited: Interplay of individual differences and repeated exposures. *Language Learning & Technology*, 29(3), 92–134.
<https://doi.org/10.64152/10125/73649>
- Xu, W., Tan, X., Wang, C., & Fu, D. (2025). Does digital multimodal composing help improve EFL students' metacognitive knowledge? *Language Learning & Technology*, 29(3), 157–181.
<https://doi.org/10.64152/10125/73651>
- Zhang, E. D., & Yu, S. (2024). *Assessing L2 digital multimodal composing competence*. Routledge.

About the Editors

Bronson Hui is an Assistant Professor of Second Language Acquisition at the University of Maryland. He teaches and conducts research in areas such as instructed second language acquisition, with a focus on the cognitive aspects of vocabulary learning and teaching, and on the effects of multimodal input (e.g., reading-while-listening).

E-mail: bhui@umd.edu

ORCID: <https://orcid.org/0000-0003-1822-3109>

Matt Kessler is an Associate Professor of Applied Linguistics at the University of South Florida, where he teaches in the MA program in Applied Linguistics and the PhD program in Linguistics and Applied Language Studies. His research focuses on issues related to computer-assisted language learning, L2 writing, and genre-based pedagogy.

E-mail: mattjkess@gmail.com

ORCID: <https://orcid.org/0000-0001-5264-0059>