*Language Teaching* (2026), 1–17
doi:10.1017/S0261444826101402

CAMBRIDGE UNIVERSITY PRESS

THINKING ALLOWED

# Generative artificial intelligence and L2 writing: A research agenda

Matt Kessler

University of South Florida, Tampa, FL, USA
Email: mattjkess@gmail.com

(Received 2 January 2026 UTC; accepted 12 June 2026 UTC)

**Abstract**

Generative artificial intelligence (GenAI) is a powerful technology that has vast potential to support various aspects of second language acquisition (SLA). Given GenAI's capabilities, it is particularly relevant for the teaching and learning of second language (L2) writing. Despite its potential, there are also clear hazards and a range of potentially negative side effects, many of which have yet to be explored. Building on existing research, in this piece, I propose a series of six future research tasks that may prove useful for further understanding the affordances and limitations of GenAI for L2 writing. These six research tasks are organized into three interrelated themes, which cover 1) learning processes and outcomes, 2) student use and interactions, and 3) teaching. For each theme, two research tasks are proposed. Each task includes a discussion of what research is needed, why it is needed, along with how scholars might investigate that research task by adopting quantitative, qualitative, or mixed methods. The goal of this piece is to provide potential research ideas for graduate students and faculty and, ultimately, to foster research–pedagogy connections involving GenAI, L2 writing, and SLA.

**Keywords:** chatbots; educational technology; foreign languages; GenAI; L2 writing

## 1. Introduction

Generative artificial intelligence (or GenAI) has been with us for some time, yet it was OpenAI's 2022 release of the chatbot ChatGPT that stirred the public's imagination, catapulting the technology into mainstream consciousness. ChatGPT itself is a form of GenAI—a large language model trained on massive amounts of data—designed to respond to human prompts by producing a wide range of outputs. In fact, this definition can be confirmed by asking ChatGPT, which responded to my prompt of "What is generative artificial intelligence?" with the following:

> Generative artificial intelligence (AI) refers to a branch of AI that focuses on creating new content or data that resembles existing data, rather than simply recognizing or classifying it. This technology uses algorithms, particularly generative models, to produce a wide range of outputs, including text, images, music, and even video. (OpenAI, 2025)

This description provided by ChatGPT highlights the broad capabilities of GenAI tools: that is, not only can they generate linguistic text, but many can also produce other modes such as sound and images (still and moving). As Crosthwaite and Sun (2025) have noted, GenAI's capacity to generate and remix content distinguishes it from other forms of AI commonly studied in applied linguistics,

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press

such as automated speech recognition tools and automated writing evaluation/scoring systems. As such, ChatGPT's novelty and generative capabilities sparked an AI-arms race among tech companies, and within three years, the digital marketplace was flooded with a plethora of GenAI tools which are now available for free or via a subscription (see Pegrum, 2025 for a related discussion of GenAI tools for teaching and learning purposes).

Due to GenAI's powerful capabilities to mimic human speech and produce diverse modes of communication, many applied linguists have found themselves dabbling in GenAI-related research; some have even completely pivoted away from their research agendas, turning their attention solely to GenAI-related matters. Within applied linguistics, scholars in the subfield of second language (L2) writing have been especially active. In small part, L2 writing researchers' interests stem from the fact that when using chatbots such as ChatGPT, Copilot, or Gemini, among others, users often prompt the chatbots via written text (although many tools permit users to dictate prompts orally as well).

More crucially, however, is that the emergence of GenAI tools has created distinct opportunities and challenges for L2 writing teachers and administrators (Barrot, 2023). In terms of opportunities, GenAI tools have the capacity to save teachers' time by supporting their lesson planning and materials creation efforts (Kohnke & Zou, 2025; Li et al., 2025; Lin et al., 2025); additionally, GenAI tools can aid teachers and their students in facilitating written corrective feedback (WCF; Muñoz et al., 2025; Pfau et al., 2023). On the other hand—as readers are undoubtedly aware—GenAI tools also have the potential to be misused for more nefarious purposes, such as cheating or plagiarism. For example, students can use chatbots to instantly complete many writing activities that are commonly assigned in novice- and intermediate-level language courses (Casal & Kessler, 2023; Goulart et al., 2025). Thus, since the act of writing itself is thinking—a form of externalized thought, reflection, and communication—many students, teachers, and researchers alike have voiced fears that GenAI's proliferation will lead to a loss of intelligence and critical thinking abilities (e.g., Guo, 2025; Xu & Tan, 2024).

Given its potential affordances and dangers, over the past few years, there has been a flurry of activity surrounding GenAI and L2 writing. Multiple recent syntheses and reviews have attempted to digest this rapidly expanding body of research and to make sense of early findings (e.g., Crosthwaite & Sun, 2025; Feng et al., 2025; Li, 2025; Luo et al., 2025). As these early reviews indicate, much of the work on GenAI and L2 writing can be divided into three primary strands, including research on: 1) WCF and assessment, 2) teachers' and students' perceptions of GenAI's utility, and 3) GenAI literacy. In sum, reviews suggest that although there are some differences between teacher and GenAI feedback (at the moment), GenAI tools appear to be effective for facilitating WCF and subsequent uptake (see Crosthwaite & Sun, 2025). Additionally, while teachers and students often report positive perceptions of their interactions with GenAI chatbots and the types of feedback they receive, there is also an underlying sense of mistrust and concerns about how using such tools might impact learners both academically and cognitively (e.g., Zou & Huang, 2023). Finally, in terms of literacy, unsurprisingly (i.e., given the novelty of GenAI tools), studies have shown that few teachers are prepared to harness GenAI, and many (if not most) will require extensive training (e.g., Moorhouse et al., 2024); relatedly, while numerous students have reported experimenting with GenAI tools, many have not (e.g., Kessler et al., 2025). Thus, questions remain about when to teach GenAI literacy, how to do so, and what the development of GenAI literacy might look like over time (a topic I return to later).

Although much has been accomplished in such a brief period, our understanding of GenAI still remains quite limited, particularly in terms of its positive and negative impacts on the teaching and learning of L2 writing. Despite how little we know, it is interesting to note that there is already a GenAI-backlash and fatigue percolating throughout the field. For instance, in recent publications (and at conferences and editorial board meetings), many journal editors have publicly lamented the skyrocketing submissions they are receiving on the topic of GenAI (e.g., Hu, 2026); this fatigue, apparently, has led to many desk rejections, not because of quality, but simply due to the topic itself. While I understand this perspective from an editor's point of view, I would also strongly argue that this

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press

is somewhat shortsighted and misses the larger point, which is that: GenAI is the most powerful and consequential technology we have seen—as both educators and as people more generally—since Apple first popularized the smartphone nearly 20 years ago. More crucially, due to GenAI's capabilities, it has the capacity to be *far* more impactful, both in positive and in negative ways, which will shape our personal and our professional lives in a myriad of unknown ways in the years to come.

On a related note, we, as L2 writing educators, know that many of our students (and colleagues) are using GenAI tools—some in basic ways, and some in increasingly complex ways. Thus, the reality is also this: GenAI is not going away, and our students are not going to stop using it. However, like any technology, GenAI is a tool. Thus, as researchers, it is now our responsibility to figure out the broad questions of: When should we use GenAI for L2 writing? And, when should we not use it (and why)?

Therefore, although many may be tiring of GenAI's recent takeover of the field, if anything, we need much more research, not less. We also need considerable replication work and overlap in terms of study/topic areas. Although replication is often misunderstood by many applied linguists and L2 writing scholars as being "low prestige" or "unoriginal" (see Mu & Matsuda, 2016, p. 202), in reality, replication work is the bedrock of good and meaningful science (see Porte & McManus, 2019).

As such, in response to this need for additional research that expands our understanding of GenAI's place and roles within L2 writing, I propose a series of research tasks in this piece, with the aim of expanding upon and clarifying aspects of scholarship that have already been published. In the next section, I describe the organization of the proposed research tasks and the themes that motivated their development.

## 2. Research tasks

In this section, I propose six research tasks that address the intersection of GenAI and L2 writing. When presenting these tasks, they are grouped thematically into three subsections. As mentioned, many studies to date have explored three major topics, one of which includes WCF and assessment. By far, WCF has been the most investigated topic within L2 writing (e.g., Crosthwaite & Sun, 2025; Feng et al., 2025). Therefore, apart from a brief mention of WCF (in research task #4), I do not propose tasks related to this topic. For those readers who are interested in GenAI and WCF, they are encouraged to see the aforementioned articles for thorough discussions of feedback-related agendas. Instead, in this piece, I propose research tasks that address other critical areas that are more broadly related to three interconnected themes, which include 1) LEARNING PROCESSES AND OUTCOMES, 2) STUDENT USE AND INTERACTIONS, and 3) TEACHING. In what follows, I begin by presenting an overview of each research theme. After introducing a theme, I then propose two research tasks that correspond to that theme. When discussing each research task, I also describe what research is needed and why, and I suggest potential study designs and methods as a means of tackling the research task under discussion. This format is then repeated (i.e., introduction to a research theme, followed by two related research tasks).

Notably, these three thematic topic areas have been selected for multiple reasons. First, these themes share overlapping and interrelated characteristics—that is, LEARNING PROCESSES AND OUTCOMES relate closely to STUDENT USE AND INTERACTIONS, which, in turn, also involve aspects of pedagogy and TEACHING. Therefore, it is possible that a task in one theme (e.g., TEACHING) might subsequently inform a task or future study in another theme (e.g., LEARNING PROCESSES AND OUTCOMES). Second, apart from being interrelated, the proposed six tasks each cover a topic where research has already begun. In this way, each task contains example studies that can be used as motivation (e.g., for conducting a related study, for replication purposes, or for other methodological inspiration) and as a point of comparison with future findings. Within each task, many of the studies were selected to showcase because they are of good quality in multiple ways (e.g., being methodologically rigorous, transparent in the reporting of findings). These example studies exemplify some of the basic pillars of study quality (see Plonsky, 2024). Third and finally, these themes and tasks were selected because

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press

they address some of the practical, day-to-day realities of being an L2 writing teacher. Thus, the findings of future studies in these areas have the capacity to inform research and practice for L2 writing instructors and teacher educators.

### 2.1. Research theme: Learning processes and outcomes

Research involving GenAI has followed a predictable trend in that, whenever a new technology or digital tool emerges, one of the very first topics that researchers often rush to investigate is stakeholders' perceptions of it. Of course, developing a thorough understanding of people's perceptions is vital. After all, if students or their teachers do not like the tool, do not find it useful, or find it challenging to use, then it is unlikely that they will continue using it for educational purposes in the future. That being said, when it comes to new educational technologies, studies of actual learning tend to come second. Thus, the first theme presented here, LEARNING PROCESSES AND OUTCOMES, involves investigating how GenAI tools can foster aspects of the learning process, as well as facilitate positive L2 learning outcomes.

As discussed, most studies involving GenAI and learning processes and outcomes to date have focused on the extent to which GenAI tools can be useful for WCF purposes. In their synthesis of 51 WCF studies, Crosthwaite and Sun (2025) note that many studies have examined how GenAI tools can lead to improvements in writing quality, as well as topics involving the accuracy of GenAI feedback, and how students engage with the feedback they receive. Relatedly, studies have also examined how teachers interact with and use GenAI for providing feedback (e.g., Han & Li, 2024). Continued research involving GenAI and WCF will be critical, as this research has many practical implications for teachers and learners in terms of formal and informal assessment. However, we also need studies that go beyond examining feedback, as learning to write extends far beyond seeking and receiving WCF.

Writing, particularly in an L2, is a complex process that involves various cognitive and social dimensions (Hyland, 2022; Williams, 2012). Additionally, depending on the stages of one's L2 proficiency, students may be engaging with writing for very different purposes. At lower levels of proficiency, learners are primarily engaging in LEARNING-TO-WRITE activities, in which they are learning basic skills such as spelling, manipulating grammatical forms, and constructing sentence- or paragraph-level discourse in highly controlled situations (see Manchón, 2011). At more advanced stages of proficiency, students are often engaging in WRITING-TO-LEARN tasks. In this sense, the writing tasks that students engage in may resemble real-life tasks, mimic disciplinary genres, or be reflective in nature, in which the process of writing is used as a mechanism for prompting students to engage in critical thinking and to learn or develop new ideas (see Leow, 2020). Due to the complexities involved in writing, research is needed that investigates a myriad of dimensions, particularly when it comes to LEARNING PROCESSES AND OUTCOMES.

**Research task 1: Investigate L2 writers' GenAI literacy and how it develops over time.**

As Keefe and Copeland (2011) note, literacy is often viewed as a basic skill that many children learn during elementary schooling; in reality though, it is a far more complex. Traditionally speaking, literacy has involved one's capacity to read and write, yet today, it is also viewed as the capacity to interpret, comprehend, and critically evaluate information from a variety of print and digital resources, as well as the capacity to manipulate and produce various modes for communicative purposes (Cope & Kalantzis, 2015; Pangrazio et al., 2020). With the development of GenAI technologies, GenAI literacy has also now emerged as a new dimension of literacy in multiple respects (Warschauer et al., 2023). For one, as more students continue to use GenAI tools, there is a difference between simply using GenAI tools and knowing how to use them ethically and productively (i.e., prompting chatbots in

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press

ways that avoid cheating and using them to support one's work and learning rather than offloading all cognitive work). Additionally, AI literacy is increasingly a desired skill beyond the classroom, as many employers are seeking individuals who can teach them how to use and integrate AI into the workplace (Ryan, 2025).

As mentioned, GenAI literacy is one of the three major topics of inquiry that has been investigated within L2 writing. With such studies, many have focused on teachers. These studies have often investigated aspects of teachers' knowledge of GenAI tools, their capacity to use them for pedagogical purposes, and more (e.g., Moorhouse, 2024; Moorhouse & Kohnke, 2024; Wang et al., 2025). Due to the novelty of GenAI as a technology as well, some L2 writing scholars have proposed their own literacy frameworks (e.g., Darvin, 2025; Ou et al., 2024; Wang & Wang, 2025). For example, Ou et al. (2024) have argued that in order to understand critical GenAI literacy, a four-part framework is required, which includes concepts such as INTERACTION WITH AI, ETHICS IN ACADEMIC WRITING, TECHNICAL VALUE AND LIMITATIONS, and SELF-LEARNING. It is notable that this framework—like other researchers' proposed frameworks—all share similar constructs. When aggregated, these frameworks often attempt to tap into key issues of literacy, such as users' knowledge of what GenAI tools can and cannot do generally, knowledge of the focal tool's functions, along with how such tools can be used ethically to support one's learning.

Although literacy-related research continues to emerge, relatively little has been published on L2 writers' GenAI literacy and how it develops over time in conjunction with other academic literacy skills (e.g., Tan, 2025; Lim & Darvin, 2026). One illustrative example is a study by Tan (2025). In her study, Tan explored how one L2 English graduate student leveraged ChatGPT to support her academic literacies development during the first year of her graduate program. Tan triangulated data such as interviews, survey responses, and logs from ChatGPT. Although the heart of Tan's article is about academic literacies development, through examining the student's use of ChatGPT over one year, it also details aspects of the participant's GenAI literacy over time. In her findings, Tan traces some of the positive aspects of her participant's literacy development, such as how the participant learned to effectively use the chatbot to support her completion of reading and writing tasks and to learn information about disciplinary conventions in her major. However, Tan also documents some of the negative impacts of chatbots on her academic literacy. For example, the participant voiced concerns that, over time, she felt she was "relying on it [ChatGPT] a little too much"; as a result, the participant provided the self-assessment of "I don't think my writing skills have really improved that much" over the year (p. 8).

Tan's study demonstrates the importance of further investigating GenAI literacy, and some of the ways in which GenAI literacy may interact with and support or hinder academic literacies. In particular, we now need studies that investigate questions such as: What do students know and/or believe about GenAI tools when they enter our classrooms? In addition, how do their initial conceptualizations of these tools, along with their practices, change over time? That is, with increased literacy, to what extent are students able to use GenAI as a scaffold or support tool rather than for cognitively offloading critical academic literacy skills? Finally, to what extent can training (e.g., class sessions or workshops) impact different facets of students' GenAI literacy? Because literacy is a multifaceted construct, researchers might choose to explore a specific aspect of literacy pertaining to one of these questions. For example, with students' conceptualizations of GenAI tools and how they change over time, researchers might focus on students' knowledge of what these tools can and cannot do, as well as how students' perceptions of the ethical uses of GenAI tools evolve over part or all of one semester (see Kessler et al., 2025 for an example survey on the ethical uses of GenAI tools for L2 writing). In terms of how students' practices evolve, researchers might look at changes in students' prompts or prompt-engineering strategies (i.e., how they query a chatbot either verbally or in writing). As both Warzel (2023) and Pegrum (2025) have noted, prompt engineering is an emergent form of thinking; therefore, prompt engineering itself is a form of literacy that needs to be both taught and understood in terms of its development over time.

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press

In terms of executing these studies, Tan (2025) once again provides a useful foundation on which to build, both for researchers who are interested in qualitative or quantitative methods. That is, because literacy itself is a complex skill which may develop slowly, collecting data at multiple time points to track that development is likely necessary. For instance, in qualitative studies, this might involve researchers collecting multiple data sources for triangulation purposes, such as interviews, reflective logs/journals, chatbot logs, writing samples, think alouds, or stimulated recalls. In terms of quantitative studies, researchers might collect one or two of these types of data as well, along with another instrument such as a survey. Surveys or questionnaires, in particular, could be implemented in a pre-/post-test design or even at multiple time points spread across a semester.

Finally, for analyzing and interpreting data, there are various options. As discussed, multiple writing researchers have proposed analytical frameworks for researching and teaching GenAI literacy (e.g., Darvin, 2025; Ou et al., 2024; Wang & Wang, 2025). Some researchers have even published studies in which they have designed and validated questionnaires that attempt to tap into constructs of AI literacy (e.g., Ng et al., 2024; Wang et al., 2025). Researchers are encouraged to explore these options listed here and, ultimately, to determine which methods and frameworks align best with their own interests and also those that meet the needs and profiles of the learners in their contexts.

**Research task 2: Investigate the influence of GenAI tools on individual differences and/or emotions.**

Apart from literacy, in applied linguistics, there has also been growing recognition of the importance of individual differences (IDs) and emotions. As research has shown, in many cases, IDs (such as anxiety, motivation, willingness to communicate, and more), like emotions (such as joy, frustration, and boredom), have the capacity to either inhibit or facilitate language learning processes and outcomes. Thus, language learners' and teachers' IDs and emotions have received a significant amount of attention during the past decade (see Dewaele & Li, 2018; Li et al., 2022; Martínez Agudo, 2018). Due to the critical roles they play in second language acquisition (SLA), we have also begun to see studies emerge that investigate IDs or emotions in relation to L2 writing and GenAI.

When it comes to GenAI and IDs, scholars have recently explored topics such as motivation (e.g., Teng, 2025) and metacognition (e.g., Mizumoto, 2023; Xu et al., 2025), among others. Teng (2025), for example, conducted a study examining how motivation developed among 261 L2 English learners in China, as students engaged either with peers or GenAI for writing support. Adopting a longitudinal design with a survey administered at three different time points over one year, Teng reported that GenAI use appeared to increase learners' motivation, particularly early on; however, learners' motivation then leveled off and plateaued over time. As such, Teng notes the capacity of GenAI tools to enhance student motivation in certain phases of the writing process.

In turning from IDs to emotions, we have also seen publications that examine the emotions that students experience as they interact with chatbots (e.g., Xie et al., 2027; Zhao et al., 2025; Zou et al., 2025b). For instance, Xie et al. (2027) conducted a case study with a group of four university-level English learners in China, as they engaged with GenAI while composing a multimodal picture book. In the study, students were required to prompt a chatbot to produce visuals that could be included in their picture books. Using data such as an observation, screen recordings, stimulated recall, and a focus group interview, the authors recount the fluctuations in students' emotions. Although the focal students occasionally experienced joy when the chatbot produced what they wanted, the students mostly experienced a range of negative emotions such as frustration and, eventually, hopelessness, as a result of continually prompting the chatbot without receiving their desired results.

Crucially, these two studies by Teng (2025) and Xie et al. (2027) showcase the important roles of IDs and emotions in SLA, as well as the potential interactions between them, including how negative emotions have the capacity to demotivate students when engaging with activities. As a result, we need additional research that explores IDs and emotions independently, as well as research that probes the

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press

potential interactions between them. For instance, we need studies that investigate questions such as: When engaging with GenAI to complete a task, what emotions (e.g., boredom, frustration, joy) do students report experiencing (and why)? How does the use of GenAI impact the ID of X (such as anxiety, identity, metacognition, motivation, or willingness to communicate)? Finally, when engaging with GenAI, to what extent does X emotion impact Y ID (or vice versa)?

Methodologically, there are multiple ways to approach these research questions. For qualitative studies, case studies are a logical choice, either for investigating the emotions of individuals or pairs/groups of students as they collaborate with GenAI tools for a given task (e.g., during a singular class activity or an activity that extends over a brief period, as exemplified by Xie et al., 2027). Additionally, emotions—like IDs—have been shown to fluctuate and evolve over time. Therefore, studies are needed in which researchers adopt more longitudinal, ethnographic approaches in the classroom. This could be achieved by tracing how individuals or pairs/groups experience specific IDs and/or emotions over a semester and the extent to which this impacts students' engagement with activities. For accomplishing this, data could be collected in the form of observations (by the researcher/teacher), field notes, audio recordings of student interactions, and semi-structured interviews with students.

In terms of quantitative designs, a survey might be employed as a means of tracking changes in an ID or emotion. Once again, the time period could be brief (e.g., before and after a singular activity), or it could extend over a lengthier period. This might involve giving a pre-/post-survey or administering a survey on three or more occasions (as demonstrated in Teng, 2025). With quantitative studies involving IDs, in particular, there are questionnaires that have been proposed for evaluating aspects of IDs such as anxiety (e.g., Horwitz et al., 1986; also see Lee & Ye, 2023), metacognition (e.g., Sun et al., 2021; Zhang & Qin, 2018), and motivation (e.g., Dörnyei & Taguchi, 2010; also see Papi, 2010). Thus, researchers might adapt and employ these questionnaires to suit their needs. With quantitative studies as well, apart from describing how GenAI use impacts learners' processes, it will be particularly insightful if researchers can link IDs or emotions to specific learning outcomes. For instance, some of the aforementioned studies attempt to link IDs such as metacognition with writing performance (e.g., Sun et al., 2021). In this regard, it would be useful for researchers to combine questionnaires with elicited examples of students' written output and performance (e.g., via a timed writing task). For analyzing these data, depending on the instruments and types of variables involved, researchers might consider conducting regressions, structural equation modeling, or other statistical analyses to examine the extent to which IDs or emotions influence or predict certain learning outcomes, including any interactions among variables.

Of course, these are but a few examples of studies that might be conducted; however, the results of such studies could shed important light on GenAI's role in mediating IDs, emotions, and L2 learning processes and outcomes.

### *2.2. Research theme: Student use and interactions*

While the first theme covered various studies that delved into LEARNING PROCESSES AND OUTCOMES, the second theme addresses STUDENT USE AND INTERACTIONS. Specifically, this theme pertains to various aspects of how learners engage with GenAI tools independently and with each other. As discussed, most of the research to date exploring how students use or engage with chatbots has focused on WCF, including how students react to the feedback that they receive and the extent to which they find that feedback to be useful. However, there is now a body of literature emerging in which scholars have investigated student-initiated uses of GenAI tools (e.g., Hoomanfard & Shamsi, 2025; Hwang et al., 2025; Kessler et al., 2025; Kohnke et al., 2025). For instance, in Kessler et al. (2025), we adopted mixed methods to investigate how 287 university students enrolled in 5 different foreign languages reported using GenAI tools for L2 writing purposes. Surprisingly, in sharp contrast to various reports about widespread student use (e.g., College Board, 2025), only 14.6% of the students in our sample

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press

reported being users of GenAI tools; many also believed GenAI use to be unethical in higher education settings, unless used for specific purposes. Still, we identified 12 distinct ways in which those 14.6% of student users reported leveraging GenAI for L2 writing purposes. Of particular note is that only 1 of the 12 uses was for WCF. For instance, students reported (and subsequently demonstrated via individual case studies) using GenAI for learning how to use words appropriately in context, generating ideas, learning collocations, creating outlines, and more.

As this study and others have demonstrated, there are a myriad of ways in which students are currently engaging with GenAI tools to support their L2 writing efforts both inside and outside of the classroom. Because of this, we now need studies that explore various aspects of student use and interactions. As will be discussed, we need an array of studies that investigate how learners interact with GenAI tools individually and collaboratively, along with how students process and react to chatbot-related outputs during those interactions.

**Research task 3: Compare individual vs. collaborative engagement with GenAI tools.**

Over the past two decades, collaborative writing (CW) has been a popular topic of inquiry within L2 writing. Not only do multiple theories of SLA and human development support its use (e.g., Interactionist Approaches, Sociocultural Theory), but there is now a deep well of literature that describes its various benefits to learners (e.g., Li & Zhang, 2023b; Storch, 2019; Zhang et al., 2021; Zhang & Plonsky, 2020). Research also suggests that many L2 writing teachers now recognize these benefits and regularly adopt CW activities in their classrooms (e.g., Kessler & Casal, 2024). Of course, despite its widespread adoption, the CW literature is also filled with complexities—that is, CW is not simply good for L2 learning whenever and however it is implemented. For instance, both students' proficiencies and group size (e.g., dyads vs. groups) must be considered prior to pairing students; additionally, the types of activities and digital tools given to learners can shape the interactions they engage in. Thus, even though CW has been researched extensively, there is still a considerable amount of research that needs to be done both on novel topics (see Li & Zhang, 2023a) and in terms of replication (see Kessler, 2025).

In addition, as GenAI tools continue to make their way into classrooms, we need studies that examine students' individual and collaborative engagement with GenAI for L2 writing purposes. This is because many students are now engaging in writing activities with GenAI as their sole writing partner or "peer." Some researchers have also suggested that in instances where it may be challenging to find human interlocutors, GenAI may be a promising alternative for students (e.g., Cheng et al., 2025; Su et al., 2023). However, many teachers are also asking their students to work collaboratively in class with both a human peer and a GenAI tool (e.g., in a pair or group activity).

For example, in Michel et al. (2025), the authors argue that teachers must find creative ways to integrate GenAI into their CW assignments since GenAI is a powerful technology that has the capacity to positively influence learning outcomes. In their study, Michel et al. explored how four dyads in an L2 German class engaged with GenAI during a CW activity, which required students to compare their own writing with GenAI-produced models; the students used a rubric with statements to help guide them through the comparison and evaluation process. Analyzing data such as audio recordings of dyads' interactions and screen recordings of students' writing and revision behaviors, the researchers provide various insights into how well the dyads collaborated, what the students discussed during their interactions, and ultimately, how students reworked their papers to try to improve the overall text quality. Michel et al.'s study is but one of multiple early attempts to explore the intersections of CW and GenAI, but it is important because it suggests the potential of GenAI to be used to foster language-related episodes (LREs) (i.e., discussions about language use). It also highlights some of the complexities of CW scholarship, including the need for more research. For instance, while some dyads engaged in numerous LREs, one dyad demonstrated a lack of interaction with the chatbot, and

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press

another higher-proficiency dyad focused more on the task itself rather than language-related matters. This study highlights the need not only for mixed methods but also for future studies that compare and contrast student and teacher perception data.

As such, more research is needed that explores several topics that have previously been the focus of computer-mediated CW studies, only this time with GenAI as the technology in focus. For instance, some key questions include: Do learners prefer engaging in GenAI-supported writing activities individually or collaboratively? (Do their preferences change based on: the writing task/the students' proficiency levels/or on other factors?) Additionally, how does individual vs. collaborative engagement with GenAI impact students' written output in terms of (lexical and syntactic) complexity, accuracy, fluency, and text quality? When engaging with GenAI-supported CW activities, how do factors such as proficiency level, group size (e.g., dyads, groups of three), or other pairing considerations influence the nature of students' interactions (e.g., the number of LREs produced and resolved)? And finally, how do higher proficiency students in graduate-level courses individually and/or collaboratively engage with GenAI tools for research writing purposes or composing other academic genres?

In terms of executing such studies, fortunately, there is an extensive body of CW literature. Therefore, there are plenty of options for all researchers, regardless of whether they are interested in quantitative, qualitative, or mixed methods. Due to the number of available options, instead of recommending specific study designs here, I encourage readers to review two key resources. The first resource is Storch's (2019) research timeline on CW, which appears in this journal. Storch's timeline is an informative account of early, influential CW studies, which includes brief descriptions of the studies' designs and findings. The second resource is Zhang et al. (2021), which is a systematic review of 113 computer-mediated CW studies. This synthesis covers key features of studies' designs, including the contexts, methods, analytical frameworks, and data analyses used in prior studies. Readers are encouraged to review these two resources—particularly Zhang et al.'s synthesis on computer-mediated CW—as many of the studies that are discussed can be modified to include GenAI as the focal technology.

**Research task 4: Explore how GenAI supports and/or hinders L2 writers of different Global Englishes or of other languages (and their varieties).**

As some of the prior tasks have suggested, with GenAI's emergence, numerous benefits have been documented; however, there also have been major concerns. Plagiarism and cheating are often the primary issues discussed across academia and by the public more broadly, yet another concern—particularly from linguists and language educators—has been how GenAI tools might contain inherent biases toward speakers of different Global Englishes and toward those who are learning non-English languages and their varieties (e.g., Jeon et al., 2024; Kang & Hirschi, 2025; Lee et al., 2025). This concern stems from the types of data that many large-language models are presently built upon, which (at the time of writing this article) often lack significant training data from non-English languages and which also reflect Standard American English (SAE) preferences; thus, as scholars have noted, many of the chatbots that are available have the potential to exhibit the same problematic biases (Bommasani et al., 2022), and there tends to be tremendous variability in performance across languages (Swinehart et al., 2025). As Kang and Hirschi (2025) further remark, "many current AI models seem to replicate, and at times, even amplify or distort social biases found in human language in unexpected ways… AI systems do not process speech with diverse characteristics equitably" (p. 70). Obviously, there are many languages spoken around the world, and there are also many diverse varieties of both English and other languages. Therefore, for those who are learning an L2 that is not English, or for those who do not speak a dominant variety, the interactions they have with chatbots

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press

have the potential to be problematic. For instance, issues may arise when learners are seeking WCF. That is, if the training data are biased, then learners may be corrected unnecessarily.

One example of this is showcased by Lee et al. (2025). In their case study, the authors explored the use of different prompts within ChatGPT, explicitly asking the chatbot to consider Global Englishes principles based on the background of the study's focal student, who was a speaker/writer of Korean English. The researchers' goal was to explore whether the chatbot could integrate a Global Englishes perspective by valuing principles like linguistic diversity and communicative effectiveness, or whether the chatbot would default to providing feedback based on SAE and native-speaker norms. In their findings, Lee et al. reported that ChatGPT often resorted to native-speaker norms, focusing on standard grammar and vocabulary. Despite explicitly being prompted to do so, the chatbot failed to consider the student's background and English variety. The authors provide specific examples of grammatical and lexical items that ChatGPT incorrectly marked as being errors and subsequently replaced with standardized forms.

The study by Lee et al. (2025) is a powerful example of the dangers of GenAI to negatively impact speakers of English around the world. It also further illustrates how limited training data (in a specific variety of English or in non-English languages) not only have the capacity to provide users with incorrect responses but also, as Lee and colleagues go on to state, to "eras[e] multilingual and multicultural aspects of the writer …" (p. 94). Due to these potential dangers, we now need future studies that explore aspects of GenAI student use and interactions, involving either a) how GenAI supports or hinders L2 learners of languages other than English (and their varieties), or b) how GenAI supports or hinders L2 writers of various Global Englishes.

These lines of inquiry are necessary since, as discussed, much of what we currently know about GenAI and L2 writing involves English as the target language, primarily from studies involving American or British English. In general, we need more studies that involve languages other than English. These studies could be on any topic pertaining to how L2 writers use or interact with GenAI chatbots, including WCF and how students process and respond to that feedback. Studies might employ original designs, or they might also be replications of existing studies but with English replaced by a different target language (e.g., To what extent do the findings of X-study hold true if English is replaced by X-language?).

For studies involving Global Englishes or varieties of other languages, we need research that explores what happens when learners of different varieties interact with chatbots. For example, questions might be posed such as: What kinds of feedback do students of X-variety receive when seeking feedback on their writing (in terms of grammar, content, style, etc.)? How do they perceive that feedback? For instance, do they perceive the feedback as being (in)correct? And to what extent do these interactions impact aspects of their emotions, identities, motivation, and/or willingness to engage with chatbots? Finally, do learners have their own strategies for prompting or engaging with chatbots when learning a specific variety (and if so, what are they)?

In terms of executing these studies, Lee et al. (2025) offer a clear template of how researchers might attempt to prompt a chatbot to consider specific criteria (e.g., linguistic diversity and communicative effectiveness) when responding to learners' or researchers' inquiries. However, it is also important to note that it is highly unlikely that students will prompt chatbots with such sophisticated, research-backed prompts when using chatbots on their own. As such, in addition to studies that explore the extent to which prompt engineering might be used to make chatbots sensitive to specific varieties, we also need more naturalistic studies involving students and how they use GenAI. For instance, quantitative studies that are conducted inside a classroom might leverage data such as chatlogs to record the types of prompts students give and the types of responses and feedback they receive, in addition to using surveys to investigate how those interactions are perceived and/or influence different factors (e.g., aspects of awareness, motivation). In terms of qualitative methods, researchers might conduct similar studies inside the classroom but by leveraging data such as observations, chatlogs, and semi-structured interviews; conversely, such studies could also be conducted outside of the

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press

classroom via screen-recordings, chatlogs, semi-structured interviews, and self-report data such as reflective journals, as a means of exploring how students engage with chatbots on their own beyond school.

Once again, there are a limited number of studies in this area. Due to the paucity of research, any contributions that attempt to speak to the intersections of GenAI, Global Englishes (or non-English languages), and L2 writing are welcomed.

### *2.3. Research theme: Teaching*

While the first two themes addressed tasks pertaining to LEARNING PROCESSES AND OUTCOMES and STUDENT USE AND INTERACTIONS, the final theme addresses TEACHING. As the name implies, this theme is related to how educators might strategically leverage GenAI tools for specific purposes. Across applied linguistics, we have seen numerous studies in which researchers have advocated for and described how GenAI can be used by teachers. For instance, researchers have discussed how GenAI tools can be used for creating reading materials (e.g., Lin et al., 2025), for developing accessible and personalized speech scoring systems (e.g., Shin et al., 2025), and even for evaluating chatbots' general capabilities for supporting teachers of languages other than English (e.g., Swinehart et al., 2025). However, as readers will note, none of these aforementioned areas are chiefly concerned with L2 writing instruction. That is, apart from leveraging GenAI for providing WCF, far fewer researchers have examined the applications and utility of GenAI for other areas of L2 writing instruction.

Because of this, we need studies that explore various aspects of teaching with GenAI in the classroom. As will be discussed in the tasks that follow, we need studies that investigate how teachers might leverage GenAI tools to support two key areas that are currently prominent within L2 writing pedagogy. These two areas involve genre-based instruction (GBI) and digital multimodal composing (DMC).

**Research task 5: Investigate GenAI's capacity to facilitate GBI and the teaching of various academic and professional genres.**

GBI is a popular approach to L2 writing, which, as Hyland (2007) explains, enables teachers to ground their pedagogies in authentic texts that students will encounter in their everyday lives. Although there are multiple approaches or schools of thought when it comes to executing GBI, regardless of the approach, Tardy et al. (2023) remark that one of the primary goals is to move instructors away from teaching prescriptive, templated forms of writing (e.g., five-paragraph essays), and instead, to turn writing instruction into something that is "dynamic, situated, goal-oriented, and responsive to readers and communities" (p. 1).

Given the aims of GBI, as a pedagogy, it is often used with L2 learners of higher proficiencies (e.g., in English for Academic/Specific Purposes [EAP/ESP] courses), when students have the linguistic capabilities to engage with academic or professional genres. Notably, when it comes to L2 academic writing and GenAI, research has shown that instructors have been busy experimenting with GenAI in EAP classrooms for teaching important writing skills such as paraphrasing (e.g., Xu & Zheng, 2025). Additionally, many L2 graduate students already appear to possess complex strategies for using GenAI to support their academic research and writing efforts (e.g., Hoomanfard & Shamsi, 2025; Zou et al., 2025a). Despite the prominence of GenAI tools in many graduate students' lives and academia more broadly, there have been few attempts to explore the applications of GenAI for GBI and for supporting the teaching and learning of different genres in the L2 writing classroom.

One notable study in this regard is Kim and Lu (2024). In their study, the authors examined the potential of using ChatGPT for automating rhetorical move-step analysis—a common type of linguistic analysis that is often taught to L2 writers in EAP and ESP courses. By experimenting with

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press

various prompts, Kim and Lu analyzed ChatGPT's accuracy in annotating research article introductions from various social science disciplines (e.g., applied linguistics, economics, political science, psychology). Through experimentation and prompt refinement, the researchers demonstrated that, in some cases, ChatGPT was able to tag the rhetorical moves with approximately 90% accuracy. As Kim and Lu subsequently note, these findings have implications not only for researchers but also for instructors and GBI. They argue that GenAI tools might be used by L2 writing teachers for genre analysis activities as a means of helping novice writers tag and analyze unfamiliar texts. Thus, future studies are needed that investigate the potential applications of GenAI tools for facilitating GBI and teaching different academic and professional genres.

While Kim and Lu's (2024) study is an excellent demonstration of the capabilities of GenAI tools for move-step analysis, future studies are needed that are situated in classroom contexts with teachers and students. Research questions might be posed, such as: To what extent do teachers find GenAI useful for facilitating GBI? To what extent can GenAI be used to facilitate rhetorical move analysis activities for X-genre, without completely offloading all of students' cognitive work (i.e., supporting learning, rather than replacing it)? By extension, do students find GenAI to be accessible and useful for such purposes? And more broadly, to what extent can GenAI be used to help foster genre knowledge and awareness (e.g., of audience, moves-steps, and/or common lexico-grammatical features)?

For executing these studies, since classroom-based studies are needed, quantitative, qualitative, or mixed methods could be adopted depending on the size of one's class and the number of sections that are available for study. For the aforementioned questions, researchers interested in quantitative or mixed methods might adopt a quasi-experimental design with one or two classrooms. Since studying and learning a new genre typically takes more than one class period (with time varying widely from genre-to-genre), multiple GenAI genre-analysis activities might be integrated into a syllabus over the span of one or multiple weeks. Studies with two sections of a class could have an experimental group who adopts GenAI for certain activities, while the comparison group engages in traditional manual analyses. Conversely, if only one teacher and section are available for study, researchers might explore a standalone activity with GenAI and evaluate the teacher's and students' experiences.

In terms of data collection and analysis, quantitative or mixed methods studies might analyze samples of students' writing before and after a treatment. Alternatively, such studies could analyze pre-/post-surveys before and after the activities to evaluate aspects such as changes in students' genre knowledge or students' perceptions of the utility and usability of GenAI for learning about the focal genre. In terms of qualitative studies, multimodal visualization reflections might be a useful source of data, particularly for understanding aspects of students' genre awareness (e.g., Kessler, 2024b; Negretti & McGrath, 2018). Finally, since teachers are key to this particular research task, data such as semi-structured interviews could be useful for gaining insights.

For analyzing these data, as mentioned, there are multiple approaches to GBI (e.g., ESP, Systemic Functional Linguistics). Although they are not mutually exclusive, those who research genre often operationalize and analyze constructs differently based on the approach being taken. Because of this, data analyses and frameworks will vary widely. Interested readers are encouraged to see resources such as Hyland and Shaw (2016) and Kessler and Polio (2024), which cover conducting genre-based research in applied linguistics from multiple perspectives.

**Research task 6: Investigate GenAI's capacity to facilitate the teaching and execution of DMC tasks.**

Apart from GBI, another popular topic during the past decade has been DMC (e.g., Jiang & Hafner, 2024; Kessler, 2024a; Li, 2022; Lim & Kessler, 2024; Zhang et al., 2023). DMC involves creating meaning by leveraging digital tools and manipulating linguistic text in addition to one or more non-linguistic mode (e.g., aural, gestural, visual, or spatial). As research has demonstrated,

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press

the use of digital, multimodal genres and activities is now ubiquitous throughout many higher education contexts (Lim & Polio, 2020). As a result, multimodal literacy is considered a key competency for achieving academic and professional success (Grapin & Llosa, 2020; New London Group, 1996).

While prior studies have explored a plethora of digital tools for completing multimodal tasks, in recent years, we have seen researchers begin to explore how GenAI can be integrated into the DMC composing process (e.g., Jiang & Lai, 2025; Kang & Yi, 2023; Tan et al., 2025; Xie et al., 2027). For example, in Jiang and Lai (2025), the authors conducted a mixed-methods study examining how GenAI use differentially affected the quality of students' DMC products. In the study, 110 L2 English learners in Hong Kong (ages 11–14) worked in groups of 3–5 over several weeks to collaboratively create a digital video documentary. These 110 students were divided into 2 main groups: those permitted to use a GenAI tool called Poe ($n = 14$ groups) and those who were not ($n = 10$). In addition to analyzing students' video documentaries, the researchers collected screen recordings of the students' composing processes, conducted observations of in-class activities, and conducted pre-/post-study focus group interviews with students. Jiang and Lai reported that the GenAI-assisted groups outperformed the non-GenAI groups in multiple areas, including total score on the grading rubric, with a large effect size. In terms of qualitative data, analyses of the focus group interviews suggested that the GenAI tool served as an important scaffold for students when they were struggling with generating ideas and structuring their videos, particularly when compared to those students who did not use GenAI.

The findings of Jiang and Lai's (2025) study are noteworthy in that they speak to one of the main challenges that students and teachers often voice when it comes to engaging in DMC. Specifically, this challenge is that high levels of digital literacy may be required to engage in certain types of DMC tasks (such as composing digital videos), which, in turn, can create various technical/usability issues that teachers need to spend time resolving. Additionally, from a student's perspective, it can take considerable time to find certain non-linguistic modes (e.g., audio, images) to integrate into their compositions. Thus, the use of GenAI in this study and others speaks to GenAI's capacity to facilitate the teaching and execution of DMC tasks. However, since this line of research is still in its infancy, additional studies are needed.

In particular, research questions might be posed such as: To what extent do teachers find GenAI useful for generating materials to be used in DMC projects? To what extent can GenAI facilitate aspects of students' composing processes when engaging in DMC (once again, as a means of supporting learning rather than completely offloading all of students' cognitive work)? Subsequently, how does GenAI use impact aspects of students' DMC products? And finally, to what extent is GenAI useful for completing X-type of DMC task when compared to Y-type of DMC task? (That is, is GenAI perceived as being more useful for composing some multimodal genres over others)?

Methodologically speaking, classroom-based studies are needed. Additionally, quantitative, qualitative, or mixed methods could be adopted, depending on the size of one's class and the number of sections that are available. For multi-section studies, Jiang and Lai's (2025) study provides a template for how students in different classes could be divided into GenAI and non-GenAI groups to examine the impact of the technology on specific product-related outcomes (e.g., composition quality, complexity, accuracy, and fluency). Apart from quantitative and mixed methods studies, qualitative studies are also needed. Any of the research questions posed above, for instance, could be investigated by adopting a combination of qualitative data such as observations, screen recordings (of students' composing processes), stimulated recalls (of the composing process), and/or interviews. Once again, data analyses will vary widely based on the questions being posed and the methods involved; however, with qualitative studies, by triangulating multiple data sources, insights could be provided into aspects of teachers' and students' perceptions, along with their actions (e.g., composing processes and decision-making when it comes to leveraging and remixing different modes).

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press

## 3. Conclusion

As some of the studies in this article demonstrate, GenAI is a promising technology that has tremendous potential to support numerous aspects of L2 teaching and learning. However, there are also apparent dangers and a host of potentially negative side effects, many of which have yet to be documented. In this piece, I have proposed six research tasks, which I believe may prove fruitful for further understanding GenAI's various affordances and limitations. Of course, beyond these six research tasks, there are many additional lines of inquiry that can (and should) be pursued. Similarly, as GenAI research continues to expand, we do not simply need more research, but we also need GOOD QUALITY research—that is, research which is methodologically rigorous, transparent, ethical, and of value to society (for more, see Plonsky, 2024). As such, I hope that, apart from examining some of the tasks that have been proposed here, researchers will continue to take great care in their work and to be creative—both in terms of the topics they investigate and in terms of the methods they use to explore such phenomena.

## References

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, *57*, 100745. https://doi.org/10.1016/j.asw.2023.100745

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R. B., Arora, S., von Arx, S., Demszky, D., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., & Brynjolfsson, E. (2022). On the opportunities and risks of foundation models. *arXiv*, arXiv:2108.07258. https://doi.org/10.48550/arXiv.2108.07258

Casal, J. E., & Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics*, *2*(3), 100068. https://doi.org/10.1016/j.rmal.2023.100068

Cheng, D., Li, M., & Lee, T. (2025). Leveraging ChatGPT for research writing: An exploration of ESL graduate students' practices. *Computers and Composition*, *76*, 102934. https://doi.org/10.1016/j.compcom.2025.102934

College Board. (2025). New research: Majority of high school students use generative AI for schoolwork. *Newsroom*. Retrieved December 27, 2025, from https://newsroom.collegeboard.org/new-research-majority-high-school-students-use-generative-ai-schoolwork

Cope, B., & Kalantzis, M. (2015). The things you do to know: An introduction to the pedagogy of multiliteracies. In B. Cope, & M. Kalantzis (Eds.), *A pedagogy of multiliteracies* (pp. 1–36). Palgrave Macmillan. https://doi.org/10.1057/9781137539724_1

Crosthwaite, P., & Sun, S. (2025). Generative AI and L2 written feedback studies: A scoping review. *RELC Journal*, 1–13. https://doi.org/10.1177/00336882251386530

Darvin, R. (2025). The need for critical digital literacies in generative AI-mediated L2 writing. *Journal of Second Language Writing*, *67*, 101186. https://doi.org/10.1016/j.jslw.2025.101186

Dewaele, J.-M., & Li, C. (2018). Editorial of the special issue "emotions in SLA." *Studies in Second Language Learning and Teaching*, *8*(1), 15–19. https://doi.org/10.14746/ssllt.2018.8.1.1

Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). Routledge.

Feng, H., Li, K., & Zhang L. J. (2025). What does AI bring to second language writing? A systematic review (2014-2024). *Language Learning & Technology*, *29*(1), 1–27. https://hdl.handle.net/10125/73629

Goulart, L., Matte, M. L., Mendoza, A., & Alvarado, L. (2025). AI or student writing? Analyzing the situational and linguistic characteristics of undergraduate student writing and AI-generated assignments. *Journal of Second Language Writing*, *66*, 101160. https://doi.org/10.1016/j.jslw.2024.101160

Grapin, S. E., & Llosa, L. (2020). Toward an integrative framework for understanding multimodal L2 writing in the content areas. *Journal of Second Language Writing*, *47*, 100711. https://doi.org/10.1016/j.jslw.2020.100711

Guo, X. (2025). Facilitator or thinking inhibitor: Understanding the role of ChatGPT-generated written corrective feedback in language learning. *Interactive Learning Environments*, *33*(5), 3526–3544. https://doi.org/10.1080/10494820.2024.2445177

Han, J., & Li, M. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System*, *126*, 103502. https://doi.org/10.1016/j.system.2024.103502

Hoomanfard, M., & Shamsi, Y. (2025). Generative AI in dissertation writing: L2 doctoral students' self-reported use, AI-giarism, and perceived training needs. *Journal of English for Academic Purposes*, *78*, 101750. https://doi.org/10.1016/j.jeap.2025.101570

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal*, *70*(2), 125–132. https://doi.org/10.2307/327317

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press

Hu, G. (2026). GenAI! GenAI? GenAI…. *Journal of English for Academic Purposes*, *80*, 101650. https://doi.org/10.1016/j.jeap.2026.101650
Hwang, H., Chang, X., & Sun, J. (2025). Generative AI is useful for second language writing, but when, why, and for how long do learners use it? *Journal of Second Language Writing*, *69*, 101230. https://doi.org/10.1016/j.jslw.2025.101230
Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, *16*(3), 148–164. https://doi.org/10.1016/j.jslw.2007.07.005
Hyland, K. (2022). *Teaching and researching writing* (4th ed.). Routledge.
Hyland, K., & Shaw, P. (Eds.) (2016). *The Routledge handbook of English for academic purposes*. Routledge.
Jeon, J., Lee, S., & Coronel-molina, S. M. (2024). Rethinking AI: Bias in speech recognition chatbots for ELT. *ELT Journal*, *78*(4), 435–445. https://doi.org/10.1093/elt/ccae035
Jiang, L., & Hafner, C. (2024). Digital multimodal composing in L2 classrooms: A research agenda. *Language Teaching*, *58*(4), 528–546. https://doi.org/10.1017/S0261444824000107
Jiang, L., & Lai, C. (2025). How did the generative artificial intelligence-assisted digital multimodal composing process facilitate the production of quality digital multimodal compositions: Toward a process-genre integrated model. *TESOL Quarterly*, *59*(1), 52–85. https://doi.org/10.1002/tesq.3390
Kang, J., & Yi, Y. (2023). Beyond ChatGPT: Multimodal generative AI for L2 writers. *Journal of Second Language Writing*, *62*, 101070. https://doi.org/10.1016/j.jslw.2023.101070
Kang, O., & Hirschi, K. (2025). Bias and stereotyping: Human and artificial intelligence (AI). *Annual Review of Applied Linguistics*, *45*, 69–84. https://doi.org/10.1017/S026719052500008X
Keefe, E. B., & Copeland, S. R. (2011). What is literacy? The power of a definition. *Research and Practice for Persons with Severe Disabilities*, *36*(3-4), 92–99. https://doi.org/10.2511/027494811800824507
Kessler, M. (2024a). *Digital multimodal composing: Connecting theory, research and practice in second language acquisition*. Multilingual Matters.
Kessler, M. (2024b). Leveraging multimodal visualization reflections for L2 writing research. *Research Methods in Applied Linguistics*, *3*(3), 100126. https://doi.org/10.1016/j.rmal.2024.100126
Kessler, M. (2025). Replication research in L2 collaborative writing: Replications of Fernández Dobao and Bikowski and Vithanage (2016). *Language Teaching*, *58*(2), 274–282. https://doi.org/10.1017/S0261444824000302
Kessler, M., & Casal, J. E. (2024). English writing instructors' use of theories, genres, and activities: A survey of teachers' beliefs and practices. *Journal of English for Academic Purposes*, *69*, 101384. https://doi.org/10.1016/j.jeap.2024.101384
Kessler, M., & Polio, C. (Eds.) (2024). *Conducting genre-based research in applied linguistics: A methodological guide*. Routledge.
Kessler, M., Rostrán Valle, J. M., Çekmegeli, K., & Farrell, S. (2025). Generative AI for learning languages other than English: L2 writers' current uses and perceptions of ethics. *Foreign Language Annals*, *58*(3), 1219–1241. https://doi.org/10.1111/flan.70005
Kim, M., & Lu, X. (2024). Exploring the potential of using ChatGPT for rhetorical move-step analysis: The impact of prompt refinement, few-shot learning, and fine-tuning. *Journal of English for Academic Purposes*, *71*, 101422. https://doi.org/10.1016/j.jeap.2024.101422
Kohnke, L., & Zou, D. (2025). The role of ChatGPT in enhancing English teaching: A paradigm shift in lesson planning and instructional practices. *Educational Technology & Society*, *28*(3), 4–20. https://doi.org/10.30191/ETS.202507_28(3).SP02
Kohnke, L., Zou, D., & Su, F. (2025). Exploring the potential of GenAI for personalized English teaching: Learners' experiences and perceptions. *Computers and Education: Artificial Intelligence*, *8*, 100371. https://doi.org/10.1016/j.caeai.2025.100371
Lee, K., & Ye, Y. (2023). Investigating the reliability of foreign language classroom anxiety scale (FLCAS): An application of generalizability theory. *Research Methods in Applied Linguistics*, *2*(1), 1000036. https://doi.org/10.1016/j.rmal.2022.100036
Lee, S., Jeon, J., McKinley, J., & Rose, H. (2025). Generative AI and English language teaching: A global Englishes perspective. *Annual Review of Applied Linguistics*, *45*, 85–108. https://doi.org/10.1017/S0267190525100184
Leow, R. P. (2020). L2 writing-to-learn: Theory, research, and a curricular approach. In R. M. Manchón (Ed.), *Writing and language learning: Advancing research agendas* (pp. 95–120). John Benjamins.
Li, M. (2022). *Researching and teaching second language writing in the digital age*. Springer.
Li, M., Belpoliti, F., Taha, G., & Zhang, M. (2025). World language teachers' practices of integrating ChatGPT into instructions: A positioning analysis. *Foreign Language Annals*, *58*(4), 999–1017. https://doi.org/10.1111/flan.70041
Li, M., & Zhang, M. (2023a). Collaborative writing in L2 classrooms: A research agenda. *Language Teaching*, *56*(1), 94–112. https://doi.org/10.1017/S0261444821000318
Li, M., & Zhang, M. (Eds.) (2023b). *L2 collaborative writing in diverse learning contexts*. John Benjamins.
Li, S. (2025). Generative AI and second language writing. *Digital Studies in Language and Literature*, *2*(1), 122–152. https://doi.org/10.1515/dsll-2025-0007
Li, S., Hiver, P., & Papi, M. (Eds.) (2022). *The Routledge handbook of second language acquisition and individual differences*. Routledge.
Lim, J., & Kessler, M. (2024). Multimodal composing and second language acquisition. *Language Teaching*, *57*(2), 183–202. https://doi.org/10.1017/S0261444823000125

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press

Lim, J., & Polio, C. (2020). Multimodal assignments in higher education: Implications for multimodal writing tasks for L2 writers. *Journal of Second Language Writing*, *47*, 100713. https://doi.org/10.1016/j.jslw.2020.100713

Lim, K. Y., & Darvin, R. (2026). Critical digital literacies, generative AI, and the negotiation of agency in human-AI interactions. *System*, *136*, 103904. https://doi.org/10.1016/j.system.2025.103904

Lin, Y., Jiang, Y., & Lei, J. (2025). Generative AI and teacher professional identity development: A case study on L2 Chinese reading materials development. *Computer Assisted Language Learning*, 1–22. https://doi.org/10.1080/09588221.2025.2592620

Luo, Y., Xia, Y., & Lu, X. (2025). A systematic review of the role of artificial intelligence in second language writing education. *Digital Studies in Language and Literature*, 1–24. https://doi.org/10.1515/dsll-2025-0001

Manchón, R. M. (Ed. ) (2011). *Learning-to-write and writing-to-learn in an additional language*. John Benjamins.

Martínez Agudo, J. D. (Ed.) (2018). *Emotions in second language teaching: Theory, research and teacher education*. Springer.

Michel, M., Bazhutkina, I., Abel, N., & Strobl, C. (2025). Collaborative writing based on generative AI models: Revision and deliberation processes in German as a foreign language. *Journal of Second Language Writing*, *67*, 101185. https://doi.org/10.1016/j.jslw.2025.101185

Mizumoto, A. (2023). Data-driven learning meets generative AI: Introducing the framework of metacognitive resource use. *Applied Corpus Linguistics*, *3*(3), 1000074. https://doi.org/10.1016/j.acorp.2023.100074

Moorhouse, B. L. (2024). Beginning and first-year language teachers' readiness for the generative AI age. *Computers and Education: Artificial Intelligence*, *6*, 100201. https://doi.org/10.1016/j.caeai.2024.100201

Moorhouse, B. L., & Kohnke, L. (2024). The effects of generative AI on initial language teacher education: The perceptions of teacher educators. *System*, *122*, 103290. https://doi.org/10.1016/j.system.2024.103290

Moorhouse, B. L., Wan, Y., Wu, C., Kohnke, L., Ho, T. Y., & Kwong, T. (2024). Developing language teachers' professional generative AI competence: An intervention study in an initial language teacher education course. *System*, *125*, 103399. https://doi.org/10.1016/j.system.2024.103399

Mu, C., & Matsuda, P. (2016). Replication in L2 writing research: Journal of Second Language Writing author's perceptions. *TESOL Quarterly*, *50*(1), 201–219. https://doi.org/10.1002/tesq.284

Muñoz, B. C. M., Nassaji, H., & Carrillo, F. I. B. (2025). ChatGPT-generated versus human direct corrective feedback on L2 writing. *System*, *134*, 103805. https://doi.org/10.1016/j.system.2025.103805

Negretti, R., & McGrath, L. (2018). Scaffolding genre knowledge and metacognition: Insights from an L2 doctoral research writing course. *Journal of Second Language Writing*, *40*, 12–31. https://doi.org/10.1016/j.jslw.2017.12.002

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, *66*(1), 60–92. https://doi.org/10.17763/haer.66.1.17370n67v22j160u

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology*, *55*(3), 1082–1104. https://doi.org/10.1111/bjet.13411

OpenAI. (2025). *ChatGPT* (Nov 29 version) [Large language model]. https://chat.openai.com/chat

Ou, A. W., Khuder, B., Franzetti, S., & Negretti, R. (2024). Conceptualising and cultivating critical GAI literacy in doctoral academic writing. *Journal of Second Language Writing*, *66*, 101156. https://doi.org/10.1016/j.jslw.2024.101156

Pangrazio, L., Godhe, A. L., & Ledesma, A. G. L. (2020). What is digital literacy? A comparative review of publications across three language contexts. *E-Learning and Digital Media*, *17*(6), 442–459. https://doi.org/10.1177/2042753020946291

Papi, M. (2010). The L2 motivational self system, L2 anxiety, and motivated behavior: A structural equation modeling approach. *System*, *38*(3), 467–479. https://doi.org/10.1016/j.system.2010.06.011

Pegrum, M. (2025). From revolution to evolution: What generative AI really means for language learning. *Language Teaching*, 1–17. https://doi.org/10.1017/S0261444825000151

Pfau, A., Polio, C., & Xu, Y. (2023). Exploring the potential of ChatGPT in assessing L2 writing accuracy for research purposes. *Research Methods in Applied Linguistics*, *2*(3), 100083. https://doi.org/10.1016/j.rmal.2023.100083

Plonsky, L. (2024). Study quality as an intellectual and ethical imperative: A proposed framework. *Annual Review of Applied Linguistics*, *44*, 4–18. https://doi.org/10.1017/S0267190524000059

Porte, G., & McManus, K. (2019). *Doing replication research in applied linguistics*. Routledge.

Ryan, R. (2025, October 28). *The no. 1 skill employers want in 2025 and most job seekers don't list it*. Forbes. https://www.forbes.com/sites/robinryan/2025/07/30/the-no-1-skill-employers-want-in-2025-and-most-job-seekers-dont-list-it/

Shin, D., Lee, J. H., & Kim, K. (2025). Exploratory study on developing and evaluating a GenAI-based speech scoring system for L2 teaching practitioners. *Innovation in Language Learning and Teaching*, 1–17. https://doi.org/10.1080/17501229.2025.2596668

Storch, N. (2019). Collaborative writing. *Language Teaching*, *52*(1), 40–59. https://doi.org/10.1017/S0261444818000320

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, *57*, 1–11. https://doi.org/10.1016/j.asw.2023.100752

Sun, Q., Zhang, L. J., & Carter, S. (2021). Investigating students' metacognitive experiences: Insights from the English as a foreign language learners' writing metacognitive experiences questionnaire. *Frontiers in Psychology*, *12*, 1–15. https://doi.org/10.3389/fpsyg.2021.744842

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press

Swinehart, N., Nguyen, P., & Yeh, E. (2025). A protocol for evaluating AI chatbots' capabilities for low-resource language teachers. *Language Learning & Technology*, *29*(1), 1–21. https://doi.org/10.64152/10125/73661

Tan, X. (2025). Academic socialization with generative AI: A longitudinal case study of an L2 graduate student's academic literacies development. *Journal of Second Language Writing*, *70*, 101263. https://doi.org/10.1016/j.jslw.2025.101263

Tan, X., Xu, W., & Wang, C. (2025). Purposeful remixing with generative AI: Constructing designer voice in multimodal composing. *Computers and Composition*, *75*, 102893. https://doi.org/10.1016/j.compcom.2024.102893

Tardy, C. M., Caplan, N. A., & Johns, A. M. (2023). *Genre explained: Frequently asked questions and answers about genre-based instruction*. University of Michigan Press.

Teng, M. F. (2025). Examining longitudinal development of writing motivation in the GenAI context: A self-determination theory perspective. *Learning and Motivation*, *91*, 102157. https://doi.org/10.1016/j.lmot.2025.102157

Wang, C., & Wang, Z. (2025). Investigating L2 writers' critical AI literacy in AI-assisted writing: An APSE model. *Journal of Second Language Writing*, *67*, 101187. https://doi.org/10.1016/j.jslw.2025.101187

Wang, Y., Derakhshan, A., & Ghiasvand, F. (2025). EFL teachers' generative artificial intelligence (GenAI) literacy: A scale development and validation study. *System*, *133*, 103791. https://doi.org/10.1016/j.system.2025.103791

Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing*, *62*, 101071. https://doi.org/10.1016/j.jslw.2023.101071

Warzel, C. (2023, February 8). The most important job skill of this century. The Atlantic. https://www.theatlantic.com/technology/archive/2023/02/openai-text-models-google-search-engine-bard-chatbot-chatgpt-prompt-writing/672991/

Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing*, *21*(4), 321–331. https://doi.org/10.1016/j.jslw.2012.09.007

Xie, L., Ou, B., & Jiang, L. (2027). Students' emotions in image-GenAI-assisted picturebook making: A multimodal analysis. In M. Kessler, J. E. Casal & F. Marino (Eds.), *Multimodal practices in higher education: Analyzing and teaching diverse genres* (pp. 103–128). Routledge.

Xu, J., & Zheng, Y. (2025). Negotiating understanding, control, and authorship: L2 learners' experiences with AI-assisted paraphrasing in academic writing. *Journal of English for Academic Purposes*, *78*, 101591. https://doi.org/10.1016/j.jeap.2025.101591

Xu, W., & Tan, X. (2024). Beyond words: L2 writing teachers' visual conceptualizations of ChatGPT in teaching and learning. *Journal of Second Language Writing*, *64*, 101110. https://doi.org/10.1016/j.jslw.2024.101110

Xu, X., Qiao, L., Cheng, N., Liu, H., & Zhao, W. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *British Journal of Educational Technology*, *56*(5), 1842–1863. https://doi.org/10.1111/bjet.13599

Zhang, L. J., & Qin, T. L. (2018). Validating a questionnaire on EFL writers' metacognitive awareness of writing strategies in multimedia environments. In A. Haukås, C. Bjørke & M. Dypedahl (Eds.), *Metacognition in language learning and teaching* (pp. 157–178). Routledge.

Zhang, M., Akoto, M., & Li, M. (2023). Digital multimodal composing in post-secondary L2 settings: A review of the empirical landscape. *Computer Assisted Language Learning*, *36*(4), 694–721. https://doi.org/10.1080/09588221.2021.1942068

Zhang, M., Gibbons, J., & Li, M. (2021). Computer-mediated collaborative writing in L2 classrooms: A systematic review. *Journal of Second Language Writing*, *54*, 100854. https://doi.org/10.1016/j.jslw.2021.100854

Zhang, M., & Plonsky, L. (2020). Collaborative writing in face-to-face settings: A substantiative and methodological review. *Journal of Second Language Writing*, *49*, 100753. https://doi.org/10.1016/j.jslw.2020.100753

Zhao, T., Zhou, S., Liang, Q., & Lian, X. (2025). Modeling learners' intention to continue using generative AI in informal language learning: The role of cognitive appraisals and emotions. *Innovation in Language Learning and Teaching*, 1–29. https://doi.org/10.1080/17501229.2025.2551120

Zou, M., & Huang, L. (2023). The impact of ChatGPT on L2 writing and expected responses: Voice from doctoral students. *Education and Information Technologies*, *29*(11), 13201–13219. https://doi.org/10.1007/s10639-023-12397-x

Zou, M., Kong, D., & Lee, I. (2025a). Doctoral student's strategy use in GAI chatbot-assisted L2 writing: An activity theory perspective. *Journal of English for Academic Purposes*, *76*, 101521. https://doi.org/10.1016/j.jeap.2025.101521

Zou, M., Reinders, H., & Amjad, F. (2025b). Understanding the potential role of GenAI-mediated informal digital learning of English (GenAI-IDLE) in the Global South: AI literacy, emotions, and willingness to communicate as outcomes. *ReCALL*, 1–19. https://doi.org/10.1017/S0958344025100360

**Matt Kessler** is Associate Professor of Applied Linguistics at the University of South Florida, where he teaches in the master's program in Applied Linguistics and the doctoral program in Linguistics and Applied Language Studies. His research explores issues related to computer-assisted language learning, second language writing, and teacher education.

**Cite this article:** Kessler, M. (2026). Generative artificial intelligence and L2 writing: A research agenda. *Language Teaching*, 1–17. https://doi.org/10.1017/S0261444826101402

https://doi.org/10.1017/S0261444826101402 Published online by Cambridge University Press