OPEN ACCESS

GLODAL Emerging Educational Technologies

*Emerging Educational Technologies*, 2026, 1, 260103
https://doi.org/10.65553/EET.260103

# *Emotions and Engagement in Generative AI-Supported L2 Writing: A Systematic Review*

Xiaoshu Xu[a]

[a]School of Foreign Studies, Wenzhou University, China

Submission date: May 1, 2026; Revision date: October 15, 2026; Online: November 20, 2026

## Abstract

Generative AI increasingly supports L2 writers in planning, drafting, revising, and responding to feedback, but its value depends partly on how learners feel and engage while using it. This systematic review synthesized 87 empirical studies of emotions and engagement in GenAI-supported L2 writing, appraised study quality, and checked reporting consistency. GenAI often increased enjoyment and reduced social-evaluative pressure, but did not reliably reduce anxiety and also introduced concerns about authorship, integrity, and the tool itself. Engagement dimensions frequently diverged: behavioral or affective engagement could increase while cognitive engagement remained unchanged or shifted toward offloading. Positive emotions and favorable appraisals did not ensure deep engagement, and negative responses could prompt either withdrawal or verification. Proficiency, AI-output quality and volume, system design, and teacher mediation shaped these patterns. Across studies, more productive engagement was associated with learners retaining responsibility for interpreting, checking, judging, and revising AI output. GenAI may therefore increase comfort and participation without preserving the evaluation required for learning.

*Keywords: Generative AI, L2 writing, emotions, engagement, control-value theory, systematic review*

*Copyright © 2026 The Authors*
*Xiaoshu Xu, Corresponding author: Lisaxu@wzu.edu.cn*

## Introduction

Generative AI now takes part in the processes through which student writers plan, draft, revise, and respond to feedback (Watson Todd, 2026), and research agendas for L2 writing place it at the center of the field (Hyland, 2026; Kessler, 2026; Khanh-Quynh & Cong-Lem, 2026). Whether this change supports writing development depends not only on the texts that learners produce with AI but also on how they feel while writing with it and how they engage with what it offers. A learner who enjoys revising with an AI tool but accepts its suggestions without understanding them may submit a better text while learning little, and a learner frustrated by output beyond their level may stop using feedback altogether. Emotions and engagement therefore stand between the use of generative AI and its effects on writing development.

Research on these questions has grown quickly, and its conclusions point in different directions. Some studies report that generative AI raises enjoyment and engagement (e.g., Jiang et al., 2026; Teng & Huang, 2026), whereas others report that learners copy AI output, offload the work of writing to the tool, or feel that improved texts are no longer their own (e.g., Nguyen et al., 2025; S. Wang et al., 2026; Yang et al., 2026a). Existing reviews of generative AI and L2 writing have mapped the field broadly, synthesized AI-generated feedback, or estimated effects on writing performance (Crosthwaite & Sun, 2026; Feng et al., 2025; Ho et al., 2026; Teng, 2026). With few exceptions (Crosthwaite & Sun, 2026; Ho et al., 2026), they have treated emotions and engagement as one strand among perceptions and outcomes, and those exceptions were limited to feedback. No review has synthesized studies in which emotions or engagement are focal constructs across the uses of generative AI in L2 writing, examined how the two relate, or weighed this evidence by its quality.

This gap leaves a practical question unanswered: under what conditions does generative AI leave L2 writers more engaged, and when does it leave them merely more comfortable? The two outcomes can look alike on a questionnaire while having different consequences for learning. This distinction is especially important because GenAI may increase comfort and observable participation while reducing the evaluative work—interpreting, checking, judging, and revising—that learners themselves perform. The present review addresses this question with a systematic review and narrative synthesis of 87 empirical studies. Guided by control-value theory (Pekrun, 2006) and by models of engagement in learning and with feedback (Ellis, 2010; Fredricks et al., 2004), it examines the emotions reported, the dimensions of engagement, the relations between them, and the conditions under which they occur. It also appraises each study's quality and checks the internal consistency of its reported statistics, so that patterns resting on sound evidence can be separated from those that do not.

## Literature review

### *Emotions in L2 Writing with Generative AI*

This review draws on control-value theory to define and interpret emotions (Pekrun, 2006). The theory traces achievement emotions, such as enjoyment, anxiety, and boredom, to two appraisals: the writer's sense of control over an activity and its outcomes, and the value the writer attaches to them. Emotions in turn shape the attention, motivation, and strategies that learners bring to a task (Pekrun et al., 2002). L2 research has added that enjoyment and anxiety are distinct dimensions rather than opposite ends of one scale (Dewaele & MacIntyre, 2014) and that classroom language learning involves a wider range of emotions than anxiety alone (Shao et al., 2019).

In generative AI-supported L2 writing, learners' capacity to evaluate fluent but sometimes inaccurate AI output (Barrot, 2023) and their sense of authorship over AI-assisted text may shape appraisals of control and value, and AI use may introduce new objects of emotion,

including the tool and the rules governing its use. The review therefore coded the object of each emotion and kept emotions, which are states, distinct from appraisals and beliefs such as self-efficacy, confidence, trust, perceived usefulness, and tool acceptance.

### *Foreign Language Anxiety in EFL Classrooms*

Engagement is defined in this review as the quality of learners' involvement in a task, with behavioral, cognitive, and affective dimensions (Fredricks et al., 2004). Behavioral engagement refers to participation and effort, cognitive engagement to depth of processing and strategy use, and affective engagement to emotional reactions during involvement. Language learning research has adopted this model and added a social dimension (Philp & Duchesne, 2016), but reviews of the field have found that definitions and measures vary widely and that self-report dominates (Hiver et al., 2024; Qian et al., 2026). The review therefore coded the dimension each study measured and the method it used rather than accepting the label "engagement" at face value.

L2 writing research has also specified engagement with respect to its object. Engagement with corrective feedback has been described in cognitive, behavioral, and affective terms (Ellis, 2010), and it differs between learners and between teacher and automated feedback (Han & Hyland, 2015; Zhang & Hyland, 2018). Two findings from automated writing evaluation research are especially relevant. Learners could act on a grammar checker's suggestions while processing them only shallowly (Koltovskaia, 2020), and trust in automated feedback shaped whether they accepted or questioned it (Ranalli, 2021). With generative AI, the object of engagement can be the tool, its feedback, or a writing task that the tool has partly completed, so the review coded the object of engagement as well as its dimension.

Generative AI also makes low cognitive engagement easier to achieve. Cognitive offloading, the use of external resources to reduce the demands of a task, can free resources for higher-level work but can also replace the processing on which learning depends (Risko & Gilbert, 2016). Outside L2 writing, students who used ChatGPT improved their essays without gaining more knowledge (Fan et al., 2025), and the quality of students' peer feedback fell once AI assistance was withdrawn (Darvishi et al., 2024). Emotional engagement is also another important issue in human-GenAI collaborative language learning (Liu & Wang, 2026). The review therefore coded offloading separately from cognitive engagement and distinguished writing produced with AI from writing produced without it.

### *The Relation Between Emotions and Engagement*

Emotions and engagement are theoretically linked but analytically distinct. Control-value theory proposes that emotions influence engagement through attention, motivation, and strategy use and that the results of engagement feed back into appraisals (Pekrun, 2006), and L2 research has argued that cognition and emotion cannot be separated in learning (Swain, 2013). The tripartite model, however, includes affective engagement as a dimension of engagement, so an emotion measured during a task can be read as either an emotion or a form of engagement (Fredricks et al., 2004). The review coded emotions measured as affective engagement as engagement findings, recorded the emotions they contained, and examined relations in both directions as a separate question.

### *Prior Reviews and the Present Review*

Reviews of generative AI and L2 writing have mapped contexts, uses, instructional designs, and language skills (Feng et al., 2025; Lee et al., 2026; Luo et al., 2025; Tang et al., 2026),

synthesized feedback generated by large language models (Crosthwaite & Sun, 2026; Ho et al., 2026), and estimated effects on writing performance that depended on studies' risk of bias (Teng, 2026). Affective and engagement outcomes recur across these reviews but are not their central object.

How emotions and engagement change, relate, and depend on conditions across the uses of generative AI in L2 writing therefore remains unresolved. The present review addresses three research questions:

RQ1: What emotions have L2 writers reported in generative AI-supported writing, what are these emotions directed at, and how do they compare with those under non-AI conditions?

RQ2: How do L2 writers engage behaviorally, cognitively, and affectively with generative AI, AI feedback, and AI-supported writing tasks?

RQ3: How are emotions and engagement related, under what learner, tool, instructional, and institutional conditions do these relationships vary, and what evidence links them to writing outcomes?

## Method

### *Design*

This systematic review with narrative synthesis is reported according to PRISMA 2020 (Page et al., 2021). A narrative synthesis was chosen over a meta-analysis because the included studies differed in design, constructs, comparison conditions, and outcomes, and most were qualitative or mixed-methods studies whose findings could not be expressed as effect sizes (Popay et al., 2006).

### *Eligibility Criteria*

A study was eligible if it (E1) was an empirical primary study with human participants; (E2) involved a generative AI tool, that is, a chatbot or writing assistant built on a large language model or feedback generated by one, rather than earlier automated writing evaluation, grammar checkers, or machine translation alone; (E3) was set in L2 writing, with learners or teachers of a second or foreign language and writing, writing feedback, or writing instruction as a focus; (E4) treated emotions or engagement as a focal construct, named in the aims or research questions or forming a main reported finding; and (E5) was an English-language journal article published online between November 30, 2022, the public release of ChatGPT, and October 5, 2026 in one of the 98 journals included in the predefined sampling frame, which was constructed from SSCI-indexed journals in Education & Educational Research, Linguistics, or Psychology, Educational. Emotions included discrete emotions and general affect, and engagement included behavioral, cognitive, and affective engagement with an AI tool, AI feedback, or a writing task. Studies that pooled AI and non-AI data, or mentioned emotions or engagement only in passing, did not meet E4. E4 was met through the aims or research questions in 79 included studies and through a main reported finding in eight.

### *Information sources and search*

The search, run on October 6, 2026, combined database searches, a journal-by-journal search within the predefined 98-journal sampling frame, and the candidate pool of the companion review. Semantic Scholar and ERIC were searched with four concept blocks combined with AND: generative AI (e.g., ChatGPT, large language model), writing (e.g., writing, essay), L2 (e.g., EFL, second language), and emotion or engagement terms (e.g., engagement, anxiety, enjoyment). Self-report terms were not required, so that qualitative studies were retrieved. The journal-by-journal search covered 98 SSCI-indexed journals in applied linguistics, L2 writing, educational technology, and related areas and retrieved every article since November 2022

with a generative AI term in its title or abstract. An OpenAlex search failed because of rate limiting, but OpenAlex records were represented in the companion pool. Search strings appear in Supplementary Material S1.

### *Selection Process*

In the database route, records were de-duplicated, and records outside the date window or the sampling frame, or lacking a generative AI, writing, L2, or emotion or engagement term in the title and abstract (the concept rule), were removed by rule. The remaining records were screened on titles and abstracts, and full texts of included and unsure records were assessed, with reasons for exclusion recorded. Title-and-abstract screening and full-text eligibility assessment were conducted by one reviewer and checked by a second reviewer. Any disagreements were resolved through discussion until consensus was reached. Because the concept rule removes any record whose abstract is missing from the source database, a second route audited all such records that fell within the date window and sampling frame and had a generative AI term in the title. Abstracts were retrieved from Semantic Scholar, OpenAlex, Crossref, or publisher websites, and records were then screened and assessed using the same review procedure as in the database route.

### *Data Items*

A structured form developed from the frameworks in the Background recorded study context, participants, design, AI tools and their roles, the writing task, and duration. For RQ1, it recorded the emotions studied, their measurement, and their object; for RQ2, engagement dimensions, objects, and data sources; and for RQ3, relations between emotions and engagement, associated conditions, and writing outcomes, including whether outcomes were measured on writing produced with or without AI. Emotions measured as affective engagement were coded as engagement findings, and appraisals such as perceived usefulness were coded as antecedents rather than emotions. Findings were condensed into evidence units, each checked against the full text. Data extraction and coding were completed by one reviewer and checked against the full texts by a second reviewer. Coding disagreements were resolved through discussion and consensus.

### *Study Quality and Reporting Consistency*

Study quality was appraised with the Mixed Methods Appraisal Tool (MMAT), which suits reviews combining qualitative, quantitative, and mixed-methods studies (Hong et al., 2018). Consistent with MMAT guidance, appraisal was interpreted at the criterion level rather than collapsed into an overall quality score. The synthesis therefore focused on design-specific methodological strengths and weaknesses rather than ranking studies by the number of criteria met. Because syntheses in applied linguistics have documented problems in statistical reporting (Plonsky, 2013), test statistics were also recomputed from reported descriptives with scripts, following the logic of GRIM and statcheck (Brown & Heathers, 2017; Nuijten et al., 2016). A study was flagged for a reporting-consistency concern when statistics used in the synthesis could not be reproduced or were internally impossible, when effect sizes and test statistics disagreed widely, or when cited references could not be matched to existing publications. Isolated errors that left conclusions unchanged were not flagged, and flags do not imply misconduct. Scripts and outputs are provided in Supplementary Material S3, and the full list of included studies appears in Supplementary Material S1; included studies cited in the text are marked with an asterisk in the references.

### *Synthesis Methods*

Following narrative synthesis guidance (Popay et al., 2006), study characteristics were tabulated, and evidence units were grouped by research question and sorted by direction and

by the conditions under which findings occurred. Control-value theory guided the interpretation of emotions, and the engagement frameworks guided that of engagement. Each synthesis claim was assigned its supporting studies, whose number, designs, and flags were computed by script. Study counts were used to describe the distribution of reported findings rather than as pooled estimates of effect or as a vote-counting rule. Quality informed the weight given to findings rather than inclusion: experimental and longitudinal findings were weighted above single-occasion self-reports, and in a sensitivity analysis every claim was re-examined with flagged studies removed.

## Results

### *Study Selection*

The searches and the audit together yielded 87 eligible studies (Figure 1). The database and journal searches identified 33,713 records, of which 176 remained after rule-based removal and 72 were sought in full text after screening. One was a duplicate report, and 61 of the 71 assessed met all criteria. The audit of the concept rule examined 1,414 records that had been removed because their abstracts were missing. Of these, 114 were screened on abstracts, 41 were assessed in full text, and 26 met all criteria. Across both routes, 25 of 112 assessed full texts were excluded, most often because emotions or engagement were not focal (16) or the study was not set in L2 writing (6).

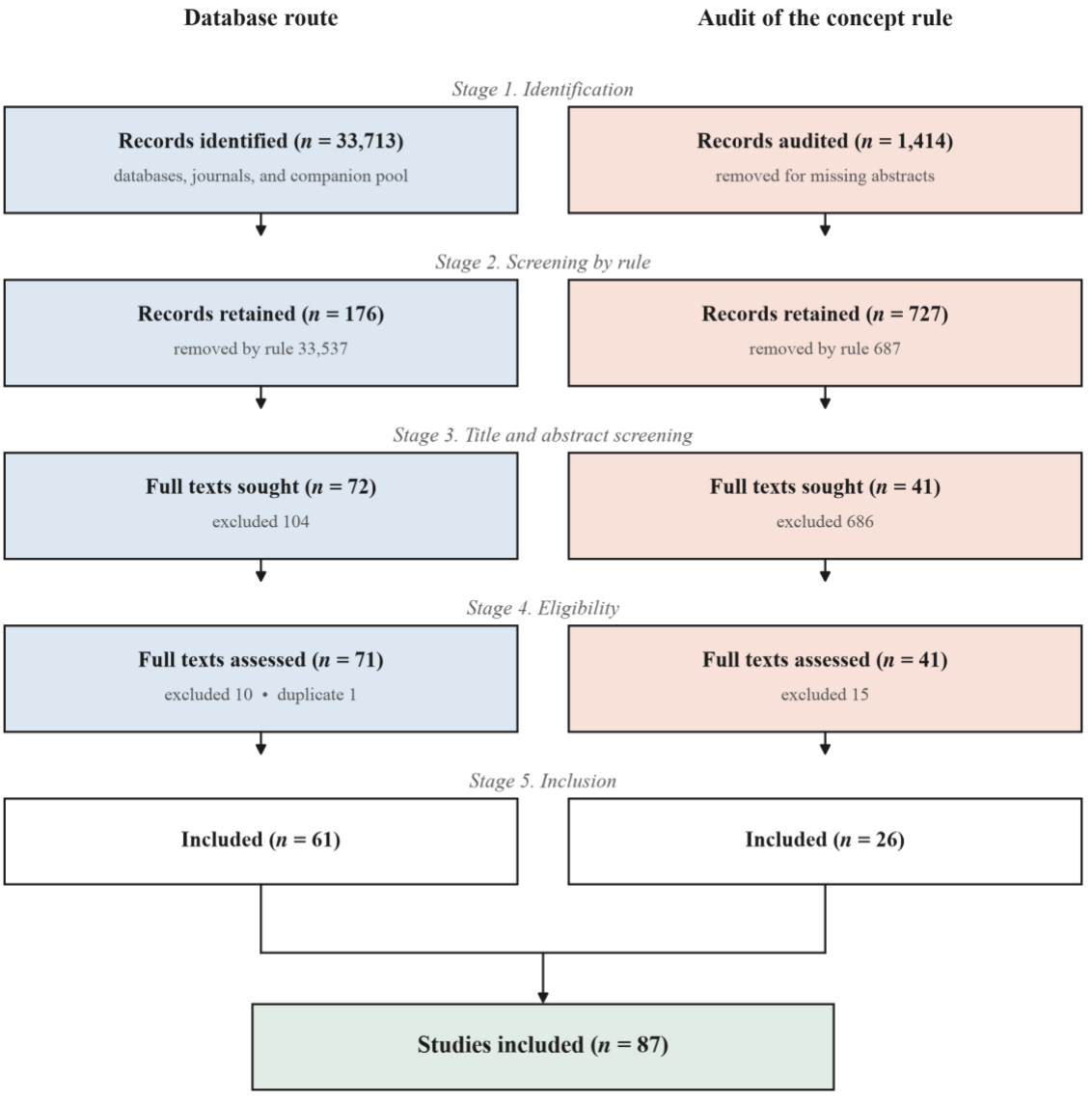



*Figure 1. PRISMA 2020 Flow Diagram of Study Selection*

*Note*. Records removed by rule were outside the date window (694), lacked a concept term (30,621), duplicated another record (978), or appeared in a journal outside the sampling frame (1,244). Full texts were excluded because emotions or engagement were not focal (database route 6, audit route 10), the study was not set in L2 writing (1, 5), no generative AI was involved (2, 0), or no participant data were reported (1, 0).

### *Characteristics and Quality of the Included Studies*

The included studies were recent, concentrated in Chinese higher education, and dominated by feedback uses of ChatGPT (Table 1). The two constructs were rarely studied together: emotions were focal in 44 studies, engagement in 58, and both in only 15. Emotions were measured mainly through interviews and rating scales, and engagement through questionnaires, interviews, revision analysis, and log or screen-recording data.

*Table 1. Characteristics of the 87 Included Studies*

| Characteristic | Category | *k* |
|---|---|---|
| **Publication year** | 2023 / 2024 / 2025 / 2026 | 1 / 6 / 28 / 52 |
| **Region** | Mainland China, Hong Kong, Macao | 56 |
| | Iran | 7 |
| | Other (13 countries or multi-country) | 24 |
| **Target language** | English | 80 |
| | German, Chinese, Japanese, French, English and Spanish | 7 |
| **Participants** | University students | 80 |
| | School students | 5 |
| | Teachers | 2 |
| **Design (MMAT category)** | Mixed methods | 45 |
| | Qualitative | 17 |
| | Quantitative non-randomized | 15 |
| | Quantitative randomized | 8 |
| | Quantitative descriptive | 2 |
| **Role of generative AI**[a] | Feedback | 72 |
| | Revision or editing | 36 |
| | Language support | 30 |
| | Idea generation | 27 |
| | Drafting or text generation | 12 |
| **Engagement dimension**[a] | Behavioral | 61 |
| | Cognitive | 57 |
| | Affective | 39 |
| **Construct focus** | Emotions focal | 44 |
| | Engagement focal | 58 |
| | Both focal | 15 |

*Note. k = number of studies.*

[a] *Studies could be coded in more than one category.*

The methodological quality of the evidence base was uneven. The most common concerns differed by design. Quantitative and mixed-methods studies were frequently limited by convenience sampling, intact-class comparisons without adequate control of baseline differences, unvalidated measures, and weak integration of qualitative and quantitative strands. Qualitative studies generally showed fewer concerns on the MMAT criteria, although the depth of sampling and analytic reporting varied across studies. Twenty-two studies, all quantitative or mixed-methods, were flagged for reporting-consistency concerns.

### *Emotions (RQ1)*

Comparative studies more often found higher enjoyment with generative AI than no difference, but the differences were small. Nine studies reported more enjoyment or positive affect with generative AI than with peer, teacher, automated, or no-AI conditions, including a randomized study in which enjoyment was higher across four writing tasks (d = 0.51; Jiang et al., 2026). Five studies found no difference or lower enjoyment, including a ten-week comparison with teacher feedback (S. Wang et al., 2026).

Comparative evidence on anxiety was mixed and rested on weaker designs. Six studies reported lower anxiety with generative AI, but three of them were flagged. Four found no difference or higher anxiety; in one, anxiety rose over ten weeks with AI feedback but stayed stable with teacher feedback (S. Wang et al., 2026). Differences concentrated in demanding tasks: in one randomized study, the AI group reported more enjoyment and less anxiety in argumentative but not narrative writing (Jiang et al., 2026).

Positive and negative emotions commonly co-occurred. Twenty-two studies reported both among the same participants: relief when the AI fixed or explained problems and frustration or confusion when its output was vague, inaccurate, or beyond learners' level. Favorable appraisals, such as greater confidence, sometimes accompanied these positive emotional responses (e.g., Koltovskaia et al., 2024; Zhao & Wang, 2026). The same tool could ease anxiety in the moment and create it later, as when reliance on AI proofreading raised anxiety about AI-free tests (W. Yang & Lin, 2025). Emotions also differed across learners and changed over time. In a study of learners of Chinese, high-proficiency learners moved from calm to happiness after ChatGPT feedback, whereas low-proficiency learners' anxiety persisted (Zhao & Wang, 2026), and in a nine-week study emotions moved from positive to negative and back as students discovered fabricated output and then used the tool selectively (Zhang & Liu, 2026).

Ten studies reported emotions directed at the AI or the rules surrounding it rather than at the writing task. These included worry about falling behind peers' AI use (Wang & Zou, 2026), guilt and fear of detection (e.g., Du & Yang, 2025; Yao & Liu, 2026), and unease about authorship when the AI rewrote students' paragraphs (Park & Ryoo, 2026). Thesis writers described distress when their self-written text was flagged by AI detectors (Ardana & Christiana, 2026). The two studies of teachers found that teachers' emotions followed their reading of students' AI use, ranging from pride when students engaged critically (Li et al., 2026) to exhaustion and anger when judging suspected AI-generated assignments (Khojasteh et al., 2025).

### *Engagement (R2)*

Behavioral engagement with AI feedback was partial and favored local suggestions. In eight of the 11 studies reporting uptake rates, learners acted on about half to two-thirds of the AI suggestions they received. Eight studies found higher uptake for local, form-focused, or specific feedback than for global or content feedback; in one case study, 67.6% of local and 34.0% of global suggestions were taken up (Ma et al., 2026). The design of the system also mattered. A system that gave diagnostic feedback without rewriting produced a median of six revision cycles, against two with free chatbot use (Park & Ryoo, 2026), although young learners' use of a feedback chatbot declined across tasks (Wolf et al., 2026).

Cognitive engagement ranged from wholesale acceptance to evaluative integration, and the range was patterned by proficiency. In 14 studies, higher-proficiency learners verified, adapted,

or rejected AI output, whereas lower-proficiency learners accepted suggestions without understanding them or pasted AI text verbatim (e.g., Lu & Jia, 2026; Rong et al., 2025). Advanced graduate writers rejected 30 of 31 inaccurate ChatGPT comments (Koltovskaia et al., 2024), whereas learners without support understood fewer than half of the suggestions they handled (Almutlaq & Alsaleh, 2026). Twelve studies reported cognitive offloading. In one planning task, 55 of 56 learners asked ChatGPT to write the whole paragraph (Nguyen et al., 2025). Five studies, however, reported a shift over weeks from wholesale acceptance toward questioning and selective use (e.g., Mohamed, 2026; Zhang & Liu, 2026).

The three dimensions of engagement did not move together. In seven studies, generative AI raised one dimension while leaving another unchanged or lower. In a class-randomized study with primary school students, behavioral and emotional engagement rose while cognitive engagement did not (Lu et al., 2026); in a randomized study of group writing, analytical talk rose while behavioral and emotional engagement did not (Hu et al., 2025). Only two studies reported gains in all three dimensions (Qi et al., 2026; Teng & Huang, 2026). Self-reported and observed engagement also diverged in two studies (Wang et al., 2026; Yang et al., 2026b).

### *Relations and Conditions (R3)*

#### *Relations Between Emotions and Engagement*

Few studies tested the relation between emotions and engagement statistically. All four that did found that anxiety was associated with lower engagement (e.g., $r = -.32$ to $-.15$; Dong, 2025), and in one of them enjoyment predicted engagement positively but more weakly than anxiety predicted it negatively (Wang & Zou, 2026).

Positive emotions and favorable appraisals did not ensure deep engagement. Five studies, including three qualitative case studies with no major concerns on the applicable MMAT criteria, reported enjoyment or satisfaction, together with favorable appraisals such as trust, alongside copying or uncritical acceptance of AI output (e.g., Cengiz et al., 2025; Yeung, 2025; Zhang & Liu, 2026). Negative emotions and unfavorable appraisals also led in two directions. Seven studies reported that frustration with feedback learners could not understand led them to skip it or avoid the tool (e.g., Takahashi & Kadowaki, 2026; Zhao & Wang, 2026), whereas six reported that distrust or disappointment prompted verification and cross-checking (e.g., Ma et al., 2026; Shi & Wang, 2026). In two studies, critical use of AI output was itself a source of positive emotion (Du & Yang, 2025; Zhong et al., 2026).

#### *Learner, Tool, Instructional, and Institutional Conditions*

Proficiency and literacy conditioned how learners felt about and used AI output. Seventeen studies reported proficiency differences in the direction described under RQ2, and two quantitative studies reported proficiency as a moderator (Abdelhadi, 2026; Sun & Jordan, 2026). Ten studies associated AI literacy, feedback literacy, or metacognition with more productive use; iterative, specific prompting yielded actionable feedback, whereas single-turn prompts yielded generic feedback and verbatim acceptance (e.g., Du & Yang, 2025; Meng et al., 2026).

The volume and accuracy of AI output conditioned negative emotions and disengagement. Fifteen studies linked long, numerous, or advanced feedback to overload and unprocessed suggestions (e.g., Liu et al., 2026), and 16 linked inaccurate or generic output to lost trust, rejection, or verification. Ten studies associated designs that restricted what the AI did or displayed its reasoning with more effortful engagement. Such designs withheld rewrites (Park & Ryoo, 2026), required learners to propose repairs first (Mohamed, 2026), or gave clues rather than corrections (Lukešová & Jennings, 2026), although most did not isolate the design feature experimentally.

The social position of the AI and the surrounding instruction also mattered. Fourteen studies reported relief because learners could ask questions and receive criticism without the evaluative pressure of a teacher or peer (e.g., Yuan & Dao, 2026; Zheng et al., 2026). Seven studies associated teacher mediation with better understanding and more positive emotions; in one randomized but flagged study, teacher explanation raised understanding of ChatGPT feedback from 57.0% to 87.7% of suggestions (Almutlaq & Alsaleh, 2026). Ten studies reported

that institutional integrity rules and detection practices generated anxiety, and detection-score thresholds amplified it, whereas checking results together with supervisors contained it (Ardana & Christiana, 2026).

### *Writing Outcomes*

Evidence on writing outcomes was limited and weaker than evidence on experience. Twelve studies compared writing produced without AI across conditions and found gains mostly in language rather than in content or organization (e.g., Lu et al., 2026; Zheng et al., 2026), but six were flagged, including the three reporting the largest gains. In five further studies, the scored text had been produced with the AI. Six studies related engagement to outcomes, two of them randomized. In the flagged one, cognitive and behavioral engagement, but not affective engagement, carried part of the treatment effect on writing growth (Sun & Jordan, 2026); in the unflagged one, engagement did not predict writing gains once group membership was modeled (Teng & Huang, 2026).

### *Sensitivity to Flagged Studies*

Most findings rested mainly on unflagged studies, but the quantitative evidence on anxiety and on writing outcomes did not (Table 2). Patterns grounded in qualitative and process data lost few or no studies when the 22 flagged studies were removed. In contrast, removing flagged studies left the evidence on anxiety balanced between lower, unchanged, and higher anxiety, and halved the evidence comparing AI-free writing across conditions.

*Table 2. Summary of Evidence for the Main Synthesis Findings*

| RQ | Finding | *k* | RCT | Other | Qual. | Flagged | *k* excl. flagged |
|---|---|---|---|---|---|---|---|
| **RQ1** | Higher enjoyment or positive affect than a comparison condition | 9 | 3 | 6 | 0 | 1 | 8 |
| **RQ1** | No difference or lower enjoyment than a comparison condition | 5 | 1 | 4 | 0 | 0 | 5 |
| **RQ1** | Lower anxiety than a comparison condition | 6 | 1 | 5 | 0 | 3 | 3 |
| **RQ1** | No difference or higher anxiety than a comparison condition | 4 | 1 | 3 | 0 | 1 | 3 |
| **RQ1** | Positive and negative emotions co-occurring | 22 | 0 | 12 | 10 | 4 | 18 |
| **RQ1** | Emotions directed at the AI or its rules | 10 | 0 | 6 | 4 | 0 | 10 |
| **RQ2** | Uptake favoring local, specific feedback | 8 | 0 | 4 | 4 | 1 | 7 |
| **RQ2** | Range from wholesale acceptance to evaluative use, patterned by proficiency | 14 | 0 | 9 | 5 | 3 | 11 |
| **RQ2** | Cognitive offloading | 12 | 0 | 9 | 3 | 2 | 10 |
| **RQ2** | Dissociation among engagement dimensions | 7 | 1 | 6 | 0 | 0 | 7 |
| **RQ3** | Anxiety associated with lower engagement (quantitative) | 4 | 0 | 4 | 0 | 2 | 2 |
| **RQ3** | Positive emotions coexisting with shallow engagement | 5 | 0 | 2 | 3 | 0 | 5 |
| **RQ3** | Negative emotions leading to withdrawal | 7 | 0 | 3 | 4 | 1 | 6 |
| **RQ3** | Negative emotions leading to verification | 6 | 0 | 3 | 3 | 2 | 4 |
| **RQ3** | Proficiency differences | 17 | 1 | 11 | 5 | 3 | 14 |
| **RQ3** | Overload from volume or complexity of AI output | 15 | 0 | 11 | 4 | 3 | 12 |
| **RQ3** | Inaccurate or generic output undermining trust | 16 | 0 | 9 | 7 | 3 | 13 |
| **RQ3** | Relief from social-evaluative pressure | 14 | 0 | 12 | 2 | 2 | 12 |
| **RQ3** | Restrictive or explanatory AI designs with | 10 | 2 | 7 | 1 | 3 | 7 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | more effortful engagement | | | | | | |
| **RQ3** | Teacher mediation | 7 | 1 | 5 | 1 | 1 | 6 |
| **RQ3** | Integrity rules and detection as sources of anxiety | 10 | 0 | 5 | 5 | 0 | 10 |
| **RQ3** | AI-free writing compared across conditions | 12 | 5 | 7 | 0 | 6 | 6 |
| **RQ3** | Engagement related to writing outcomes | 6 | 2 | 4 | 0 | 4 | 2 |

*Note. k = number of studies; RCT = randomized studies; Other = non-randomized, descriptive, and mixed-methods studies; Qual. = qualitative studies; Flagged = studies with reporting-consistency concerns (see Method). Study lists are in Supplementary Material S3. Study counts indicate the distribution of reported findings and should not be interpreted as pooled effect estimates or as a vote-counting rule.*

## Discussion

This review synthesized 87 studies of emotions and engagement in generative AI-supported L2 writing and reached three main conclusions. Generative AI changed not only how much enjoyment or anxiety L2 writers reported but also what their emotions were about. Engagement did not rise as a whole, and cognitive engagement was the dimension most likely to stay flat or give way to offloading. More productive forms of AI-supported engagement tended to involve learners, rather than the tool, interpreting, checking, and revising AI output. The sections below interpret these findings through the frameworks introduced in the Background.

### *Generative AI Extends the Objects of Emotion in L2 Writing*

The emotional findings support the prediction of control-value theory that emotions follow appraisals of control and value (Pekrun, 2006), and they show how generative AI reshapes both appraisals in L2 writing. Control depended on the match between the learner and the output rather than on the tool. Help on demand and the AI's position outside human evaluation raised learners' sense of control, whereas output that learners could not understand or check lowered it, especially at lower proficiency. Value depended on ownership. Improvement achieved with AI did not always feel like the learner's own, so AI assistance could raise the quality of a text while lowering the pride attached to it.

The theory's distinction between activity and outcome emotions does not, by itself, fully capture the range of emotion objects identified in this synthesis. Guilt about reliance, worry about detection, and unease about authorship have the AI and its governance as their objects. One included study proposed adding appraisals of the tool to the theory (Zhong et al., 2026); the present synthesis supports this proposal and adds institutional rules as a third object, which generated anxiety for students and teachers alike. The uneven evidence on enjoyment and anxiety also fits the view that the two are distinct dimensions (Dewaele & MacIntyre, 2014). A tool can make writing more pleasant without making its outcome less uncertain, and uncertainty about outcomes is what the theory places at the root of anxiety.

### *Engagement Dimensions Decouple*

The dissociation among engagement dimensions is the most consequential finding of this review. In the two strongest designs in the corpus, behavioral and emotional engagement rose without cognitive engagement, or cognitive talk rose without behavioral or emotional engagement. Earlier research on automated writing evaluation found that learners could act on a grammar checker's suggestions while processing them only shallowly (Koltovskaia, 2020). The present synthesis shows how generative AI extends this concern from accepting individual suggestions to offloading larger portions of the writing process, as when nearly every learner in

one planning task asked the AI to write the whole paragraph. In L2 writing, the evidence that AI assistance can improve products without improving learning (Fan et al., 2025) and can weaken performance once withdrawn (Darvishi et al., 2024) thus takes the form of offloading during revision and planning (Risko & Gilbert, 2016).

The decoupling has a methodological consequence. Behavioral participation, the most easily observed and most often measured dimension, may be a less informative indicator of cognitive engagement in generative AI-supported writing, because accepting a suggestion need not involve evaluating it. A questionnaire item about revising after feedback cannot distinguish a learner who revised by reasoning from one who pasted the AI's revision. This sharpens concerns about inconsistent definitions and reliance on self-report in engagement research (Hiver et al., 2024; Qian et al., 2026), and it is consistent with the divergence between self-reported and observed engagement in the present corpus.

The relation between emotions and engagement points the same way. Control-value theory predicts that positive activating emotions support engagement (Pekrun, 2006), yet positive emotions and favorable appraisals, including enjoyment, satisfaction, and trust, coexisted with copying and uncritical acceptance. One possible explanation lies in what the emotion or appraisal concerns. When enjoyment arose from the AI's output rather than the learner's own problem solving, it may have sustained use of the tool rather than processing of the task. Trust, as an appraisal of the tool or its feedback, worked similarly. Earlier research found that trust shaped whether learners accepted automated feedback (Ranalli, 2021); here, trust in fluent output went with acceptance without verification, whereas distrust prompted scrutiny. Emotional valence alone therefore provides limited insight into engagement unless it is considered alongside learners' appraisals and their capacity to evaluate and act on AI output.

### *Productive Support Leaves the Evaluative Work to the Learner*

The conditions identified for RQ3 suggest that AI support was more productive when learners could do the evaluative work themselves. Proficiency, AI literacy, and metacognition allowed learners to verify, reject, and adapt output, whereas low proficiency led to acceptance, skipping, or confusion. Research on teacher and automated feedback found that engagement depended on individual and contextual factors (Han & Hyland, 2015; Zhang & Hyland, 2018); generative AI raises the threshold, because learners must judge not only whether a suggestion is correct but whether the AI understood their intent. Restrictive or explanatory AI designs and teacher mediation may have helped learners meet this higher threshold by supporting rather than replacing their evaluative work. This interpretation extends the view that feedback literacy involves appreciating, judging, and acting on feedback and managing the emotions it provokes (Carless & Boud, 2018).

Taken together, these findings suggest that generative AI complicates the interpretation of familiar indicators of affect, engagement, and writing development. Enjoyment may reflect a more comfortable writing experience rather than deeper learning; feedback uptake may reflect acceptance rather than reasoning; and high scores for AI-assisted texts may reflect the tool's contribution rather than independent writing development. One potentially important distinction is whether learners retain responsibility for evaluating AI output. Although this interpretation has not been tested as a causal mechanism, it shifts attention from how much learners use AI to what cognitive work they continue to perform.

The evidence behind these conclusions is uneven. Patterns grounded in qualitative and process data were robust to the removal of flagged studies, whereas comparative evidence on anxiety and evidence on writing outcomes were not. Claims that generative AI reduces writing anxiety or improves AI-free writing should therefore be treated as provisional.

## Implications for Writing Instruction

Three principles for instruction follow from the synthesis, although they are inferences from patterns across studies rather than tested interventions. First, teachers can preserve the evaluative work of writing by asking AI tools for diagnosis, questions, or clues rather than rewrites, and by asking learners to propose their own revisions before consulting the AI. Second, teachers can match AI output to learners' processing capacity by limiting the volume of feedback and explaining it for lower-proficiency learners, since overload led to skipped feedback and frustration. Third, teachers can assess engagement through learners' decisions and their reasons, such as which suggestions they accepted or rejected and why, rather than through the amount of AI use or learners' satisfaction with it, and can use occasional writing without AI to gauge independent performance. Suspected misuse is better handled through dialogue than through detection scores alone, given the anxiety that detection practices generated.

## Conclusion

This systematic review examined emotions and engagement in 87 empirical studies of generative AI-supported L2 writing. The evidence suggests that AI can make writing more enjoyable and reduce social-evaluative pressure without necessarily increasing cognitive engagement, while its effects on anxiety and independent writing development remain less certain. By separating what L2 writers feel, which dimensions of engagement change, and who performs the evaluative work, the review shows that GenAI may increase comfort and observable participation without preserving the evaluative work required for learning. AI-supported writing should therefore be judged not only by learners' satisfaction or the amount of AI use, but also by whether learners continue to interpret, check, judge, and revise AI output themselves.

This review has several limitations. The restriction to English-language articles within a predefined 98-journal SSCI sampling frame limits the comprehensiveness and generalizability of the review, and the concentration of studies in Chinese higher education further limits generalization. The use of single-session designs in many studies leaves longer-term patterns unclear. Although an audit recovered 26 studies that rule-based filtering had removed, eligible studies without abstracts or without generative AI terms in their titles may still have been missed. Heterogeneity in designs and measures precluded quantitative pooling.

Future research should combine self-reports with process data, examine emotions and engagement longitudinally and in both directions, and assess writing development on texts produced without AI. Teachers' emotions, examined in only two studies, also warrant attention. Most importantly, future studies should examine whether learners' evaluative work helps explain the relationship between AI support and subsequent independent writing development.

## Acknowledgement

We gratefully acknowledge the anonymous reviewers for their diligent efforts in reviewing this study.

## Declaration of Interest Statement

There is no potential conflict of interest.

## References

References marked with an asterisk indicate studies included in the review that are cited in the text. The full list of the 87 included studies is provided in Supplementary Material S1.

*Abdelhadi, A. (2026). From pen to prompt: Costs of AI overreliance on students' cognitive endurance, self-efficacy and social engagement in writing tasks. *Education and Information Technologies*. Advance online publication. https://doi.org/10.1007/s10639-026-14131-9

*Almutlaq, S. A., & Alsaleh, Z. M. (2026). Does metalinguistic explanation of teacher and ChatGPT feedback improve EFL learners' writing quality and engagement? Findings from an intervention study. *Assessment & Evaluation in Higher Education*, *51*(3), 591–609. https://doi.org/10.1080/02602938.2025.2586124

*Ardana, S. G., & Christiana, M. (2026). AI detectors, student anxiety, and authorial alienation: A qualitative study of affective control. *Teaching in Higher Education*, *31*(6), 1081–1100. https://doi.org/10.1080/13562517.2026.2643825

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

Brown, N. J. L., & Heathers, J. A. J. (2017). The GRIM test. *Social Psychological and Personality Science*, *8*(4), 363–369. https://doi.org/10.1177/1948550616673876

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

*Cengiz, B. C., Bilki, Z., Ataş, A. H., & Celik, B. (2025). Exploring second language writers' engagement with ChatGPT feedback: Revision behaviors and perceptions. *System*, *134*, Article 103837. https://doi.org/10.1016/j.system.2025.103837

Crosthwaite, P., & Sun, S. (2026). Generative AI and L2 written feedback studies: A scoping review. *RELC Journal*, *57*(1), 207–219. https://doi.org/10.1177/00336882251386530

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education*, *210*, Article 104967. https://doi.org/10.1016/j.compedu.2023.104967

Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching*, *4*(2), 237–274. https://doi.org/10.14746/ssllt.2014.4.2.5

*Dong, L. (2025). Development and validation of a multidimensional ChatGPT feedback engagement scale in second language writing. *System*, *135*, Article 103875. https://doi.org/10.1016/j.system.2025.103875

*Du, C., & Yang, Y. (2025). Doctoral students' use of generative AI and their engagement in English for research publication purposes writing: An activity theory perspective. *System*, *134*, Article 103828. https://doi.org/10.1016/j.system.2025.103828

Ellis, R. (2010). Epilogue. *Studies in Second Language Acquisition*, *32*(2), 335–349. https://doi.org/10.1017/S0272263109990544

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, *56*(2), 489–530. https://doi.org/10.1111/bjet.13544

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://doi.org/10.64152/10125/73629

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing*, *30*, 31–44. https://doi.org/10.1016/j.jslw.2015.08.002

Hiver, P., Al-Hoorie, A. H., Vitta, J. P., & Wu, J. (2024). Engagement in language learning: A systematic review of 20 years of research methods and definitions. *Language Teaching Research*, *28*(1), 201–230. https://doi.org/10.1177/13621688211001289

Ho, N. N. Y., Rappa, N. A., & Wishart, J. (2026). Using LLM chatbots for feedback in L2 English academic writing: A scoping review. *Innovation in Language Learning and Teaching*. Advance online publication. https://doi.org/10.1080/17501229.2026.2698704

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, *34*(4), 285–291. https://doi.org/10.3233/EFI-180221

*Hu, W., Tian, J., & Li, Y. (2025). Enhancing student engagement in online collaborative writing through a generative AI-based conversational agent. *The Internet and Higher Education*, *65*, Article 100979. https://doi.org/10.1016/j.iheduc.2024.100979

Hyland, K. (2026). Writing in the AI era: Rethinking writing, research and teaching. *Journal of Second Language Writing*, *72*, Article 101302. https://doi.org/10.1016/j.jslw.2026.101302

*Jiang, Z., Huang, Z., & Yang, N. (2026). When AI support shapes affect in L2 writing: Task-specific emotions across genre-based writing tasks. *System*, *140*, Article 104041. https://doi.org/10.1016/j.system.2026.104041

Kessler, M. (2026). Generative artificial intelligence and L2 writing: A research agenda. *Language Teaching*. Advance online publication. https://doi.org/10.1017/s0261444826101402

Khanh-Quynh, T. T., & Cong-Lem, N. (2026). The new writing partner? EFL students' real-time use of GenAI tools in academic composing. *Applied Language Sciences, 2,* 260107. https://doi.org/10.65553/ALS.260107

*Khojasteh, L., Shokrpour, N., & Moslehi, S. (2025). Decision-making in decoding AI-generated content: Emotional dynamics and pedagogical strategies in English for specific purposes education. *Teaching and Teacher Education*, *157*, Article 104952. https://doi.org/10.1016/j.tate.2025.104952

Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, *44*, Article 100450. https://doi.org/10.1016/j.asw.2020.100450

*Koltovskaia, S., Rahmati, P., & Saeli, H. (2024). Graduate students' use of ChatGPT for academic text revision: Behavioral, cognitive, and affective engagement. *Journal of Second Language Writing*, *65*, Article 101130. https://doi.org/10.1016/j.jslw.2024.101130

Lee, S., Choe, H., Zou, D., & Jeon, J. (2026). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments*, *34*(1), 335–359. https://doi.org/10.1080/10494820.2025.2498537

*Li, Z., Zhang, Y., & Jiang, L. (2026). Exploring the emotional experiences of implementing GenAI-assisted digital multimodal composing: An autobiographic self-study. *Innovation in Language Learning and Teaching*. Advance online publication.

https://doi.org/10.1080/17501229.2026.2666826

*Liu, J., Shi, Z., & Li, W. (2026). Developing students' feedback literacy in disciplinary academic writing through generative artificial intelligence. *Assessing Writing*, *68*, Article 101030. https://doi.org/10.1016/j.asw.2026.101030

Liu, S., & Wang, Y. (2026). Validating and profiling emotional engagement in human–GenAI collaborative language learning. *Emerging Educational Technologies, 1,* 260104. https://doi.org/10.65553/EET.260104

*Lu, Q., Wang, M., Yao, Y., Zhu, X., Xiao, L., & Yin, H. (2026). Exploring the impact of a GenAI-supported writing program on primary students' writing motivation, engagement, and performance: A mixed methods study. *Learning and Instruction*, *106*, Article 102451. https://doi.org/10.1016/j.learninstruc.2026.102451

*Lu, Y., & Jia, C. (2026). Comparative analysis of high and low performers' use of generative AI in EFL academic writing: Behavioral patterns and perceptions. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2026.2665351

*Lukešová, A., & Jennings, P. J. (2026). Clue before correction: ChatGPT-enhanced strategy for promoting autonomous and reflective language learning. *Innovation in Language Learning and Teaching*. Advance online publication. https://doi.org/10.1080/17501229.2025.2612532

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*, *2*(2), 302–325. https://doi.org/10.1515/dsll-2025-0001

*Ma, X., Zhang, C., & Lee, I. (2026). L2 learners' engagement with AI-generated feedback on writing. *Assessing Writing*, *68*, Article 101020. https://doi.org/10.1016/j.asw.2026.101020

*Meng, Y., Cui, Y., Geng, X., & Cao, Y. (2026). Exploring feedback literacy in L2 students' academic writing: Insights from their reported engagement with GenAI for typical revision activities. *The Asia-Pacific Education Researcher*, *35*(3), 559–570. https://doi.org/10.1007/s40299-025-01051-4

*Mohamed, R. A. A. (2026). Designing traceable, prompt-scaffolded generative AI feedback to support EFL grammar accuracy and writing complexity: Evidence from male Saudi undergraduates. *Interactive Learning Environments*, *34*(8), 6031–6056. https://doi.org/10.1080/10494820.2026.2649549

*Nguyen, L. Q., Le, H. V., & Nguyen, P. T. (2025). A mixed-methods study on the use of ChatGPT in the pre-writing stage: EFL learners' utilization patterns, affective engagement, and writing performance. *Education and Information Technologies*, *30*(8), 10511–10534. https://doi.org/10.1007/s10639-024-13231-8

Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, *48*(4), 1205–1226. https://doi.org/10.3758/s13428-015-0664-2

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, Article n71. https://doi.org/10.1136/bmj.n71

*Park, Y., & Ryoo, J. (2026). Promoting revision in L2 writing: A mixed-methods study on the role of specialized AI feedback agents. *Language Learning & Technology*, *30*(1), 1–20. https://doi.org/10.64152/10125/73705

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions,

corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, *36*, 50–72. https://doi.org/10.1017/S0267190515000094

Plonsky, L. (2013). Study quality in SLA. *Studies in Second Language Acquisition*, *35*(4), 655–687. https://doi.org/10.1017/s0272263113000399

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). *Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC Methods Programme*. Lancaster University.

*Qi, Z., Liu, Y., Liu, W., & Xu, W. (2026). How does AI match students' proficiency levels to provide support? Student-driven AI-mediated scaffolding in EFL writing instruction. *System*, *142*, Article 104119. https://doi.org/10.1016/j.system.2026.104119

Qian, J., Li, D., & Li, S. (2026). Student engagement with written feedback in second language writing: A systematic methodological review. *Applied Linguistics Review*, *17*(3), 1389–1416. https://doi.org/10.1515/applirev-2025-0264

Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing*, *52*, Article 100816. https://doi.org/10.1016/j.jslw.2021.100816

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

*Rong, M., Yao, Y., Li, Q., & Chen, X. (2025). Exploring student engagement with artificial intelligence-guided chatbot feedback in EFL writing: Interactions and revisions. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2025.2539979

Shao, K., Pekrun, R., & Nicholson, L. J. (2019). Emotions in classroom language learning: What can we learn from achievement emotion research? *System*, *86*, Article 102121. https://doi.org/10.1016/j.system.2019.102121

*Shi, S., & Wang, Y. (2026). The effects of AI-powered genre-based writing instruction on low-proficiency EFL writers' ideal L2 writing self, writing enjoyment and L2 writing performance: A quasi-experimental mixed-methods study. *European Journal of Education*, *61*(3), Article e70791. https://doi.org/10.1111/ejed.70791

*Sun, L., & Jordan, C. (2026). Dynamic feedback mechanisms in technology-enhanced language learning: Examining learner engagement and second language writing outcomes. *Language Teaching Research*. Advance online publication. https://doi.org/10.1177/13621688261451408

Swain, M. (2013). The inseparability of cognition and emotion in second language learning. *Language Teaching*, *46*(2), 195–207. https://doi.org/10.1017/S0261444811000486

*Takahashi, J., & Kadowaki, Y. (2026). Exploring student engagement with ChatGPT feedback: A classroom study of JFL learners. *Foreign Language Annals*. Advance online publication. https://doi.org/10.1111/flan.70072

Tang, W., Wang, Z., & Yan, G. (2026). What are we missing in AI-supported L2 writing research? A systematic review from 2014 to 2024. *Journal of Computer Assisted Learning*, *42*(3), Article e70262. https://doi.org/10.1002/jcal.70262

Teng, M. F. (2026). Effectiveness of GenAI in enhancing writing performance: A meta-analysis. *Asian-Pacific Journal of Second and Foreign Language Education*, *11*(1), Article 70. https://doi.org/10.1186/s40862-026-00444-2

*Teng, M. F., & Huang, J. (2026). Assessing ChatGPT feedback for EFL learners' engagement and writing performance. *International Journal of Applied Linguistics*. Advance online publication. https://doi.org/10.1111/ijal.70155

*Wang, S., Wang, Y., & Solhi, M. (2026). Teacher-student interpersonal behaviours and GenAI feedback on students' achievement emotions in EFL writing: An intervention study. *European Journal of Education*, *61*(1), Article e70395. https://doi.org/10.1111/ejed.70395

*Wang, W., & Zou, M. (2026). Unveiling the role of L2 writing feedback engagement in ChatGPT-mediated language learning environment: Teacher support, L2 motivation, and emotions as predictors. *European Journal of Education*, *61*(2), Article e70584. https://doi.org/10.1111/ejed.70584

*Wang, Y., Hu, W., & Guo, Q. (2026). GPT versus teacher feedback on English writing: An analysis of feedback characteristics and student engagement. *System*, *138*, Article 103967. https://doi.org/10.1016/j.system.2025.103967

Watson Todd, R. (2026). Using generative AI to revise articles: How much is too much? *Applied Language Sciences, 2,* 260108. https://doi.org/10.65553/ALS.260108

*Wolf, M. K., Suhan, M., Deane, P., Sova, L., & Alderman, J. (2026). Young L2 students' use of an AI-assisted writing assessment and feedback tool: An exploratory study in multiple settings. *Assessing Writing*, *69*, Article 101083. https://doi.org/10.1016/j.asw.2026.101083

*Yang, F., Li, K., & Qian, L. (2026a). Student engagement with AI feedback: Evidence from ethnic minority students in China. *International Journal of Applied Linguistics*, *36*(3), 1938–1949. https://doi.org/10.1111/ijal.70066

*Yang, F., Qian, L., & Li, K. (2026b). GAI vs. AWE in writing feedback: A comparison from student engagement perspective. *Educational Psychology*. Advance online publication. https://doi.org/10.1080/01443410.2026.2659713

*Yang, W., & Lin, C. (2025). Translanguaging with generative AI in EFL writing: Students' practices and perceptions. *Journal of Second Language Writing*, *67*, Article 101181. https://doi.org/10.1016/j.jslw.2025.101181

*Yao, G., & Liu, Z. (2026). ChatGPT feedback and emotional engagement in L2 writing: A control-value theory perspective using q-methodology. *Assessing Writing*, *68*, Article 101045. https://doi.org/10.1016/j.asw.2026.101045

*Yeung, S. (2025). University students' engagement with generative AI-supported automated writing evaluation (AWE) feedback. *Journal of Second Language Writing*, *68*, Article 101203. https://doi.org/10.1016/j.jslw.2025.101203

*Yuan, S., & Dao, P. (2026). ESL tertiary learners' engagement in ChatGPT-assisted second language writing. *Language Awareness*. Advance online publication. https://doi.org/10.1080/09658416.2026.2652333

Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, *36*, 90–102. https://doi.org/10.1016/j.asw.2018.02.004

*Zhang, Z., & Liu, T. (2026). Exploring the dynamics of learners' engagement with generative artificial intelligence in second language argumentative writing. *International Journal of Applied Linguistics*. Advance online publication. https://doi.org/10.1111/ijal.70217

*Zhao, X., & Wang, D. (2026). The impact of ChatGPT's feedback on L2 Chinese learners' writing outcome, confidence, and emotions: A mixed-method quasi-experimental study. *Assessing Writing*, *68*, Article 101027. https://doi.org/10.1016/j.asw.2026.101027

*Zheng, Y., Chen, D., & Li, Y. (2026). 'From resistance to expressive drive': How a domain-specific LLM enhances L2 writing outcomes through guided feedback. *Innovation in Language Learning and Teaching*. Advance online publication. https://doi.org/10.1080/17501229.2026.2687097

*Zhong, Y., Tan, S., Lei, W., & Wang, X. (2026). Unpacking GAI-induced academic emotions: L2 learners' emotional responses to GAI-assisted grammatical complexity development in L2 writing. *System*, *139*, Article 104031. https://doi.org/10.1016/j.system.2026.104031

## Author Bios

**Xiaoshu Xu** is professor at Wenzhou University, PhD supervisor at Macao Polytechnic University and Stamford International University (Thailand), researcher of Macau Global Education Research and Innovation Center at City University of Macau. As Executive Editor-in-Chief of JETI, her research focuses on personalized learning environments, teacher professional development, and EFL teaching practices.

E-mail: Lisaxu@wzu.edu.cn

ORCiD: https://orcid.org/0000-0002-0667-4511