frontiers in Education

TYPE Review
PUBLISHED 19 June 2026
DOI 10.3389/feduc.2026.1799346

Check for updates

OPEN ACCESS

EDITED BY
Raona Williams,
Ministry of Education, United Arab Emirates

REVIEWED BY
Daniel Andrade Girón,
National University Jose Faustino Sanchez Carrion, Peru
Asrani Lit,
University of Malaysia Sarawak, Malaysia

*CORRESPONDENCE
Anastasiya A. Lipnevich
a.lipnevich@gmail.com

RECEIVED 29 January 2026
REVISED 04 May 2026
ACCEPTED 19 May 2026
PUBLISHED 19 June 2026

CITATION
Lipnevich AA, Taranto CA, DeLuca C, Nukpetsi E, Rickey N, Hopfenbeck T, Carter E and McGrane J (2026) Mapping the landscape of AI-driven feedback in education: a scoping review.
Front. Educ. 11:1799346.
doi: 10.3389/feduc.2026.1799346

Mapping the landscape of AI-driven feedback in education: a scoping review

Anastasiya A. Lipnevich[1]*, Carmen A. Taranto[2], Christopher DeLuca[3], Ephraim Nukpetsi[3], Nathan Rickey[3], Therese Hopfenbeck[4], Emma Carter[4] and Joshua McGrane[4]

[1]National Board of Medical Examiners, New York, NY, United States, [2]Graduate Center, City University of New York, New York, NY, United States, [3]Queen's University, Kingston, ON, Canada, [4]University of Melbourne, Melbourne, VIC, Australia

Introduction: Artificial intelligence (AI) is increasingly used in educational contexts to support learning, assessment, and instructional decision-making, particularly through automated feedback. While AI-generated feedback offers scalability, the rapid growth of this work has produced a fragmented empirical literature that warrants a scoping review.

Method: We conducted a scoping review following the PRISMA guidelines. A systematic search of ProQuest Dissertations & Theses Global and Web of Science yielded 32,580 records after deduplication, which were screened using ASReview, an AI-assisted screening tool. Following full-text review, 104 empirical studies published between 2008 and 2024 met inclusion criteria. Data were extracted on study characteristics, methodological approaches, AI tools, demographics, and key findings across K–12 and higher education settings.

Results: The reviewed literature was dominated by experimental ($n = 49$) and mixed-methods designs ($n = 44$), frequently incorporating content analysis to evaluate the nature and quality of AI-generated feedback. A substantial proportion of studies ($n = 55$) originated from Asian countries, with China ($n = 30$) as the leading contributor, followed by the US ($n = 15$). ChatGPT emerged as the most commonly examined AI tool ($n = 41$), followed by bespoke AI systems ($n = 22$). Findings revealed considerable variability in students' and educators' perceptions of AI-generated feedback; however, learning outcome evidence was more nuanced, with teacher feedback often remaining superior and hybrid approaches combining AI with human oversight consistently outperforming AI-only conditions. Critical gaps included underrepresentation of K-12 populations, low-income countries, and neurodivergent learners, as well as limited attention to ethical concerns such as algorithmic bias and data privacy.

Conclusion: This review provides a historical snapshot of the AI feedback literature at a moment of rapid expansion. The variability in findings suggests that the effectiveness of AI-generated feedback is strongly mediated by implementation context, task characteristics, and user factors. As newer studies continue to emerge, these findings should be interpreted as a foundation for ongoing synthesis efforts. The results highlight the complexity of integrating AI feedback into educational practice and underscore the need for more context-sensitive interpretations, methodologically rigorous designs, and attention to equity and ethics in future research.

KEYWORDS
AI in education, ChatGPT, educational feedback, scoping review, technology-enhanced learning

Introduction

Artificial Intelligence (AI) has rapidly transformed numerous sectors, and education is no exception, as evident in different position papers and government reports (Forsström et al., 2025; Grassini, 2023; OECD, 2026, Ofqual, 2024; UNESCO, 2025,). In recent years, AI-powered tools have gained significant traction in educational settings, supporting various instructional tasks from content delivery to assessment. Among these applications, AI-driven feedback to learners has emerged as a particularly prominent area of development and implementation (Guo and Wang, 2023; Villagrán et al., 2024; Yan and Zhang, 2024). This focus on feedback is understandable given its foundational role in learning. Feedback is a cornerstone of effective learning, offering students insights into their performance, guiding improvement, and fostering self-regulation (Guo and Wei, 2019; Hattie, 2023; Hattie and Timperley, 2007; Van der Kleij and Lipnevich, 2022). Traditionally, teachers (and to some extent, peers) have been responsible for providing feedback to learners. Yet, the increasing demands on educators, coupled with advancements in AI, have paved the way for automated feedback systems that can help mitigate challenges such as teacher workload, variability in feedback quality, large class sizes, and the need for timely, personalized feedback (Morales-Chan et al., 2024; Naz and Robertson, 2024; Tsai et al., 2024). These systems promise scalability, consistency, and personalized learning experiences, making them an attractive alternative or supplement to human feedback. Nonetheless, as the adoption of AI-driven feedback tools escalates and generates a substantial body of scholarly literature, it is essential to systematically examine their effectiveness, inherent limitations, and the broader implications they hold for pedagogical practices and learning outcomes.

Despite the growing body of research on AI feedback in education, the field remains fragmented. Studies vary widely in their focus, ranging from the technical development of AI algorithms to the pedagogical impact of automated feedback on student outcomes. While some research highlights the potential of AI to enhance learning through immediate, detailed, and adaptive feedback (Almegren, Mahdi, Hazaea, Ali, and Almegren, 2024; Chen et al, 2026, Colliot et al., 2024; Xiao and Zhi, 2023), other studies raise concerns about the quality, fairness, and interpretability of AI-generated feedback (Harati et al., 2021; Krange et al., 2023; Lu, 2019; Shi and Aryadoust, 2024). Other reviews and studies have raised concerns around the ethical use of AI, or lack of ethical proandcedures, and the need for facilitating students' critical thinking as part of the assessment and learning processes (Hopfenbeck et al., 2023, Yuyang et al., 2025).

Moreover, the methodological approaches used to investigate AI feedback differ significantly, ranging from qualitative, quantitative, and mixed methodologies, making it challenging to draw generalizable conclusions. For instance, some studies employ controlled experiments to measure learning gains (e.g., Chiu et al., 2024; Hutt et al., 2024), whereas others rely on qualitative methods to explore student, and to a lesser extent, instructor perceptions (e.g., Almogren et al., 2024; Alshammri, 2024; ElSayary, 2023; Lin et al. 2024). This diversity in focus and methodology underscores the need for a scoping review to synthesize existing findings and identify gaps in the literature.

In this paper, we aim to address this gap by conducting a scoping review of studies on AI feedback in education. Our review will summarize key findings and critically examine the methodologies employed in these studies. By doing so, we seek to provide a comprehensive understanding of the current state of research, including the strengths and limitations of AI feedback systems, their impact on student learning, and the methodological rigor of the studies that have shaped this field. Furthermore, we will explore the implications of these findings for educators, policymakers, and developers of AI tools, offering recommendations for future research and practice. A further contribution of this paper lies in its methodological demonstration of how AI-assisted tools, including ASReview and Rayyan, can facilitate large-scale evidence synthesis. The initial search yielded over 32,000 articles, which we systematically narrowed to 104 studies for final inclusion (ASReview LAB developers, 2024; Ouzzani et al., 2016). Given the rapid growth of literature on AI in education, future reviews in this area will likely face even larger initial yields, making such AI-assisted screening tools increasingly essential. The accessibility and user-friendliness of these platforms in managing our review process suggest their potential applicability to other research domains that generate substantial volumes of literature, offering a scalable approach to evidence synthesis across diverse fields.

Whereas other syntheses and meta-analyses related to AI and feedback have been undertaken, these efforts have been more narrowly focused on specific learner outcomes compared to the current review. This includes work by Ba et al. (2025) whose systematic review involving 129 papers examined the mechanisms and effectiveness of AI-assisted feedback in education on learner outcomes, actions and perceptions via a theory-informed approach guided by the Self-System Model of Motivational Development. Durak and Onan (2025) conducted a systematic review including 91 articles primarily focused on learning outcomes associated with AI-based feedback systems (e.g., motivation, self-regulation, collaborative learning processes), as well benefits and challenges of these systems (e.g., real-time personalized feedback, insufficient contextual sensitivity). Another recent example is Kaliisa et al.'s (2026) meta-analysis of 41 studies comparing the impacts of AI-generated and human feedback on learner's academic performance, feedback perception, and learning dispositions such as motivation and attitude. As demonstrated by these examples, existing reviews have been predominantly quantitative in their focus and approach as well as outcome-oriented. The current paper, however, offers a broader understanding into the trends, methodologies, participant profiles, and geographic representation of studies focused on AI and feedback, therefore illuminating important gaps, which we hope prove informative for future research in the field.

In undertaking this review, we seek to contribute to a deeper understanding of AI feedback in education. By systematically analyzing both the outcomes and methods of existing studies, we aim to identify best practices, highlight areas for improvement, and provide a foundation for future research. As AI continues to advance and integrate into educational systems, it is crucial to ensure that these tools are not only technologically sophisticated but, more importantly, pedagogically effective and scientifically sound. Beyond ethical responsibility, including minimizing algorithmic bias, addressing privacy concerns, and promoting transparency, we must also prioritize the validity and explainability of AI-generated feedback. Students and educators need to understand how AI systems arrive at their assessments, and these systems must demonstrate reliable, consistent performance across diverse contexts. This review represents a step toward that goal, offering insights that can inform the design, implementation, and evaluation of AI feedback systems in education.

Method

The format for conducting the review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 checklist (Page et al., 2020). We developed a standardized data extraction form (see Figure 1 and Supplementary Table S2). Following deduplication performed in R, the data were compiled into an Excel file. This format allowed each paper to be listed as a separate row, with column headers used to capture relevant information for the review. Specifically, columns were created to document participant demographics, research questions, research method, data collection method, data analysis method, and any additional details pertinent to managing the large volume of papers. This structured approach ensured that key components required for the scoping review were consistently extracted for each study. For instance, categorizing data collection methods enabled quantification of how many studies employed surveys, interviews, focus groups, or other approaches. Similarly, this method facilitated the enumeration of papers by characteristics such as country of publication (e.g., China or the United States). This systematic approach to data extraction enabled consistent categorization across all included papers, allowing for quantification of patterns by region, method, and data collection type, and supporting comparative analysis during synthesis.

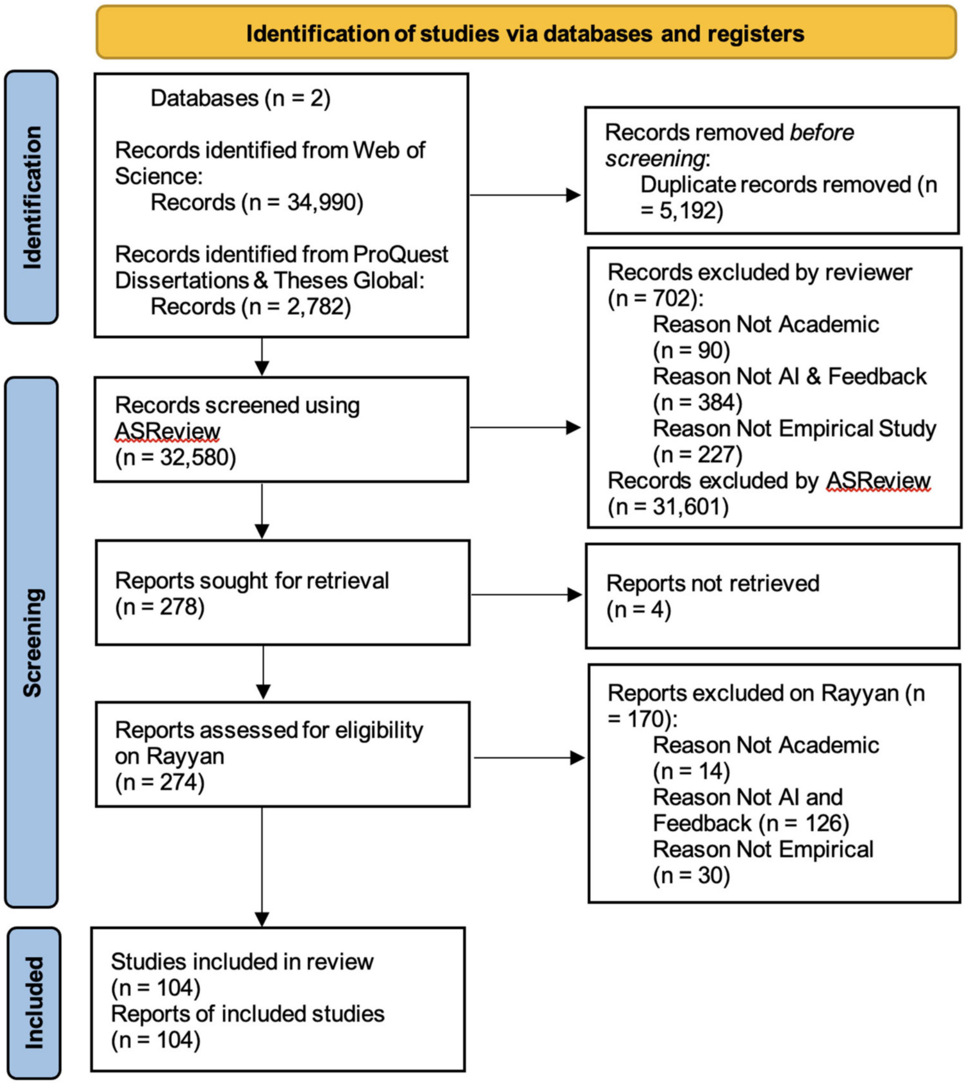

FIGURE 1
PRISMA flowchart.

Eligibility criteria

The variables of interest were AI usage in educational contexts and feedback, in education. We specified the following criteria for eligibility for our screening process:

Criterion 1: Research utilizing empirical designs, including quantitative, qualitative, or mixed methods.

Criterion 2: Studies conducted within a formal academic setting, including students from kindergarten through higher education.

Criterion 3: Studies published in English.

Criterion 4: Research focusing on educational assessment and AI, including machine learning and deep learning.

Search method

We searched the literature in the following databases: ProQuest Dissertations & Theses Global and Web of Science. In Web of Science, we utilized the structures listed in Table 1 for the search. In these databases, we conducted the search filtering results by title, abstract, and keywords. We used the following combination of terms for search in ProQuest: title(ai OR artificial intelligence OR chatGPT) AND title(feedback OR feedback processes OR comments OR review OR assessment). We used the following combination of terms for search in Web of Science: [("artificial intelligence" OR "AI" OR "machine learning" OR "intelligent tutoring system*" OR "ChatGPT") AND (feedback OR "edu* feedback" OR "performance feedback" OR "formative feedback" OR "automat* feedback" OR "personaliz* feedback" OR comment* OR review*)).

The initial search with the Boolean phrases began on July 31, 2024, and ended on September 24. We used the Boolean phrases as depicted in Table 1. These Boolean phrases were chosen after several iterations and refinements to ensure the review returned relevant literature.

TABLE 1 Search syntax, date of searches, and number of results returned for each database used in scoping review.

Database	Date of search	Results	Search syntax
ProQuest Dissertations & Theses Global	31-Jul-2024	75 (Title, English only)	title(ai OR artificial intelligence OR chatGPT) AND title(feedback OR feedback processes OR comments OR review OR assessment)
Web of Science	31-Jul-2024	4,554 (Title, English only)	(ai OR artificial intelligence OR chatGPT) AND (feedback OR "edu* feedback" OR "performance feedback" OR "formative feedback" OR "automat* feedback" OR "personaliz* feedback" OR comment* OR review*)

Selection

After all references were retrieved from the database searches ($n = 37{,}682$), duplicates were excluded using R (R Core Team, 2020) and Zotero. All identified studies were then uploaded into ASReview ($n = 32{,}580$). ASReview is an open-source, machine learning-based application designed for conducting scoping reviews that uses active learning techniques and AI to process large volumes of textual data. It ranks studies by relevance by analyzing text from titles and abstracts, considering prior decisions and selections made by human reviewers within the program regarding study inclusion and exclusion. ASReview was configured utilizing the Naive Bayes model, with feature extraction set to TF-IDF, and balance and query strategies set to dynamic resampling and max uncertainty, respectively. Utilizing its algorithm, ASReview presented the studies it deemed most relevant for classification in order of relevance. By ranking the studies in order of relevance, ASReview accelerates the screening process by gradually narrowing down the relevant studies to a point when most remaining studies are irrelevant, reducing the need to manually screen all remaining records. This approach requires the use of a stopping rule to determine when to cease manual screening. Our stopping rule for ceasing the screening process combined performance and time considerations. Thus, we established as a stopping rule that we would end the ASReview screening process when either: (a) 3% (~1,000 titles/abstracts) of our total dataset was manually reviewed, or (b) no new inclusions were found in 100 consecutive screened articles. We note that no formally validated stopping criterion currently exists in the literature, with reported thresholds in published reviews ranging from 3% to 25%; the application of a stopping rule is accordingly acknowledged as a limitation of the present review, as a more exhaustive screening process may have identified additional relevant studies.

Studies flagged as relevant by ASReview were carried forward to full-text review. Excluded abstracts were not manually verified. One researcher conducted full-text screening of all studies suggested by ASReview, consulting with the research team when questions of relevance arose.

Prior to full coding, a pilot exercise was conducted on 10 randomly selected studies to refine the coding framework and ensure consistent application of the operational definitions. Following the pilot, one researcher coded the full sample. To establish reliability, a second researcher independently coded 20% of the sample, selected randomly. Inter-rater agreement was calculated using Cohen's kappa, yielding $\kappa = 0.87$, indicating strong agreement. Discrepancies were resolved through discussion between the two researchers until consensus was reached.

The title and abstract screening yielded 278 studies that we sought for full-text retrieval. Full texts of relevant studies were obtained and evaluated for eligibility using Rayyan, an AI-powered review management software. While we sought all resources, we were unable to source four papers, bringing our sample to 274 documents for full-text review. Utilizing Rayyan, we excluded 170 papers: 14 were not in academic contexts, 126 were not researching AI and education together, and 30 studies were not empirical. The full-text screening left

104 papers, published between 2008 and 2024, in our sample for scoping review. More details on search and coding are provided in Supplementary Tables S2–S4 of the Supplementary materials.

Results

Through our scoping review of empirical evidence on AI feedback in education, we synthesized key findings pertaining to the outcomes of AI feedback and the methodologies used to study these outcomes. Here, we first summarize the reviewed literature (i.e., AI tools examined, contexts of studies, research methods) before mapping the key findings in relation to (a) perceptions and impact of AI feedback, (b) speed of AI feedback, (c) AI feedback in English Foreign Language (EFL) contexts, (d) AI feedback and peers, (e) new AI feedback tools, and (f) expressed concerns and ethical considerations.

AI tools examined

Table 2 presents the AI platforms examined in the reviewed studies and their respective frequencies. ChatGPT emerged as the most frequently investigated AI platform ($n = 41$), followed by bespoke AI systems ($n = 22$). We define bespoke AI as purpose-built tools and engines developed by researchers specifically for their studies to generate and deliver customized feedback. Table 3 provides a detailed breakdown of the ChatGPT versions represented in our corpus. Nine studies conducted comparative analyses of multiple AI platforms, while 19 did not specify a particular system. Studies that did not designate a specific AI platform investigated the technology more broadly; for instance, Zheng et al. (2023) explored generative AI's impact on student performance and motivation through semi-structured interviews and pre/post mathematical motivation assessments.

TABLE 2 AI of interest.

AI program	Quantity of studies
	n
ChatGPT	41
Bespoke AI	22
Unspecified "AI"	19
Multiple AI programs	8
Other	
Deep learning	2
PigAI	2
Coach Me	1
ASLHelp	1
DL-ALS	1
Eva	1
Grammarly	1
BERT	1
Python—automated feedback	1
Ripple	1
Technology acceptance model 3	1
WordTune	1
Total	104

Contexts of studies

Table 4 displays the geographical distribution of the reviewed studies and their respective frequencies. Asia and North America were the most prominently represented regions. China and the United States emerged as the leading contributors, with 30 studies conducted in China and 15 in the United States. Only four studies originated from Latin American countries, specifically Mexico, Ecuador, and Chile. Beyond China, 25 additional studies were conducted across diverse Asian contexts, including India, Japan, Taiwan, and Saudi Arabia. African representation was minimal, with only two studies based in Algeria and Ethiopia. European contributions comprised seven studies, including research conducted in France and Germany.

The reviewed studies focused on tertiary students from a wide range of educational disciplines. Table 5 shows the distribution of studies according to participants' sociodemographic characteristics. Most of these studies, 76, focused on higher education, including graduate, medical school, and non-matriculated university populations.

Research methodologies

Table 6 presents the distribution of research methodologies employed in the reviewed studies. In total, 44 studies utilized mixed-methods designs, out of these, 20 adopted qualitative approaches, and 40 employed quantitative approaches. Additionally, 49 studies used experimental designs, of which 23 were controlled experiments investigating the effects of feedback and AI interventions. Twenty-five studies incorporated pre- and post-test analyses to examine changes following feedback source interventions.

Quasi-experimental designs, involving manipulation of feedback conditions without random assignment, were employed in 28 studies, while correlational research, examining

TABLE 3 ChatGPT version.

ChatGPT version	Quantity of studies
	n
Not specified	18
ChatGPT 3.5	14
GPT-4	5
ChatGPT-3	4
GPT-3.5-turbo model	2
ChatGPT v.4 turbo (preview 1106)	1
GPT-3.5 and GPT-4	1
Total	45*

TABLE 4 Regions included in studies.

Region	Quantity of studies *n*
Asia	55
China	30
Taiwan	8
Saudi Arabia	3
Turkey	3
Japan	2
India	2
Iran	2
Asia-Pacific region	1
Indonesia	1
Qatar	1
Singapore	1
United Arab Emirates	1
Europe	7
Germany	2
France	1
Finland	1
Greece	1
Netherlands	1
Norway	1
North & South America	20
United States	15
Ecuador	2
Canada	1
Chile	1
Mexico	1
Australia	4
Africa	2
Algeria	1
Ethiopia	1
Multiple	1
Unspecified	13

TABLE 5 Sociodemographic characteristics of participants.

Sociodemographic characteristics of participants	Quantity of Studies *n*
Highest educational level	
Primary school	5
Middle school	3
High school	5
Undergraduate degree	69
Postgraduate degree	16
Student	93
Educator	12
AI itself	2
Other educational context	5
Unspecified	1

Some studies included multiple groups in their sample; hence, the totals exceed 104.

TABLE 6 Research methodology.

Research methodology	Quantity of studies *n*
Data collection	
Pre- and post-tests[a]	38
Questionnaires	38
Interviews	30
Content analysis	53
Learning analytics	26
Surveys	46
Focus groups	8
Observations	5
Secondary data analysis	3
Case study	1
Research method	
Experiment	49
Mixed-methods	44
Qualitative	35
Quasi-experimental	28
Controlled experiment	23
Observation	5

Some studies used multiple sources of data & methods; hence, the totals exceed 104.
[a]Includes one study that only had a post-test.

relationships among variables, was represented in only one study. Researchers utilized diverse data collection methods, including Likert-scale and frequency surveys ($n = 46$), questionnaires with open-ended questions ($n = 38$), interviews ($n = 30$), focus groups ($n = 8$), and various other approaches.

The studies used various sources of experimental evidence to answer each of their research questions. Twenty-four studies looked at perceptions of AI feedback, which included surveys and interviews of students and educators. Below, we discuss this in further detail. Another twenty-four studies investigated differences in academic performance, including grades.

Perceptions and impact of AI feedback

The first key theme emerging from the reviewed studies is perceptions of AI feedback. The perception of AI feedback has sparked a lot of research interest: twenty-four studies examined the perspectives of students and instructors. Of these, four explored perceptions from the educators' point of view (ElSayary, 2023; Guo and Wang, 2023; Kong and Yang, 2024; Pishtari et al., 2024) and 13 from higher education students. Importantly, student and educator perceptions suggest that AI feedback may be indistinguishable from teacher feedback. For instance, Almogren et al. (2024) conducted a study with four instructors and 458 graduate and undergraduate students and found that students rated ChatGPT-generated feedback messages as useful which was highly correlated with acceptance of ChatGPT for educational purposes. In another study, instructors only considered a significant modification or rewriting of GPT-generated feedback messages about 30% of the time, signifying a reasonably high degree of agreement with GPT's feedback (Wan and Chen, 2024). Additionally, many students report that AI is increasingly personalized to their needs (Alshammri, 2024; Ghafouri et al., 2024; Xiao and Zhi, 2023). Together, these studies suggest that AI-generated feedback is generally viewed positively by educators and students.

Some studies went beyond perceptions of AI feedback's usefulness to examine its impact. Studies found positive effect sizes on student learning outcomes when AI feedback was incorporated. For example, Chiu et al. (2024) investigated the impact of AI feedback on art students' performance in an artwork appreciation course at a university. In this study, participants received support from what researchers coined conventional technology-supported art learning (CT-AL) (control group) compared with a deep-learning-based art learning system (DL-ALS) (experimental group). Students demonstrated high levels of learning achievement, technology acceptance, and learning attitudes in the experimental group as compared to the control group (Chiu et al., 2024). These findings suggest that the AI-supported DL-ALS enhanced various aspects of students' art learning experience and outcomes compared to conventional technology-supported approaches. Other studies found that AI was most effective when paired with the teacher's feedback (Pahi et al., 2024). In fact, several cases demonstrated that teacher feedback remains superior to AI feedback (e.g., Steiss et al., 2024).

Wang et al. (2024b) aimed to investigate the integration of experiential learning in Virtual Reality (VR) environments and compare the effects of traditional feedback vs. ChatGPT feedback on students' reflective thinking, learning motivation, cognitive levels, and Artificial Intelligence of Things (AIoT) physical hands-on abilities. ChatGPT feedback significantly improved learners' motivation and reflective thinking and showed considerably higher abilities in Hardware Operation Tasks (HOT) and Hardware Programming Tasks (HPT) (Wang et al., 2024b). These findings emphasize the potential of AI technology in fostering deep learning, elevating the learning experience, and improving learning outcomes in VR-based educational settings.

Two studies examined the impacts of AI feedback on learning outcomes in game-based environments. Li (2023) used causal modeling to examine wrong answers marked as right (false positives, FPs) and right answers marked as wrong (false negatives, FNs) in a computer science class at an undergraduate university in Illinois. When students wrote a correct answer, erroneously marking them incorrect (FN) led to increased engagement compared to marking them correct because participants did not pay attention to the feedback marked as right, plus they did not want to "[admit] [their] answer was wrong once marked right" (Li, 2023). Interestingly, Xu et al. (2024) aimed to develop and evaluate the effectiveness of a digital game-based AI chatbot system for teaching information technology (IT) courses. This study concludes that integrating AI chatbots with digital game-based learning strategies can enhance students' learning outcomes, higher-order thinking skills, motivation, and engagement in IT courses (Xu et al., 2024). This was the only study that examined AI and educational feedback in a game-based context.

Academic performance

Twenty-four papers examined students' academic performance. Of these, nine studies conducted quasi-experiments using a pre- and post-test design. Two studies collected pre- and post-test scores along with interviews (Chen et al., 2022; Hsia et al., 2023). Chen et al. (2022) tested students aged 10–11 on their EFL learning assignments after receiving feedback in one of two conditions that both incorporated automatic feedback, differing only in the AI component. Both groups showed improvement in English-speaking skills and reduced English-speaking learning anxiety.

Ouyang et al. (2023) collected pre- and post-data from an online engineering class in China conducted on the DingTalk platform. They recorded online group discussions, personal reflections on the utility of feedback, collaboration, and overall experience, and performance data, including group and independent write-up files. They also collected 105 collaborative discussion transcripts from DingTalk, which allowed students to discuss with one another. The experimental group applied an AI-integrated model to students' learning analytics to generate feedback, which was also accessed via DingTalk. The control group only received feedback from instructors. While the findings show that the experimental group had higher academic performance scores, the authors also note higher social engagement, as evidenced by increased interaction density. Moreover, this study stands out as a multimodal data experiment that offered unique insights into collaborative learning (Ouyang et al., 2023).

Speed of AI feedback

The speed and immediacy of AI-generated feedback emerged as a critical factor in several studies. Researchers developing AI systems for educational contexts paid particular attention to temporal dimensions of feedback delivery (Colliot et al., 2024; Fidan and Gencel, 2022; Kim et al., 2022; Shi and Deng, 2024; Villegas-Ch et al., 2024). Shi and Deng (2024) investigated the relationship between feedback delay and knowledge contribution behaviors, identifying potential moderators of this relationship. Knowledge contribution behaviors are done with intention of sharing some form of knowledge to others, like taking the time to answer a question in an online question and answer (Q&A)

community (Wang et al., 2022). Their findings revealed that a brief delay in communication (i.e., approximately one to three seconds) optimized participant performance. Interestingly, performance improvements were more pronounced when participants were explicitly informed that they were interacting with AI (Shi and Deng, 2024). Beyond mere speed, the temporal flexibility afforded by AI feedback empowered students to exercise greater autonomy over their time and learning processes (Kim et al., 2022). Across multiple studies, feedback timing correlated positively with perceived usefulness and academic performance gains (Colliot et al., 2024; Fidan and Gencel, 2022; Kim et al., 2022; Shi and Deng, 2024; Villegas-Ch et al., 2024). Future research should systematically measure temporal aspects of both emerging and established AI feedback systems.

AI feedback in language learning contexts

Several studies on automated writing evaluation in EFL settings merit particular attention. Of the 104 studies reviewed, 34 examined AI-generated feedback within English as a Foreign Language (EFL) contexts, representing the largest category in our sample. EFL learners face the dual challenge of mastering a new language while preserving cultural elements in their writing (Werdiningsih et al., 2024). In their qualitative study, Werdiningsih et al. (2024) identified three main themes from student interviews regarding AI use in EFL coursework: linguistic assistance, idea generation and structuring, and proofreading with confidence enhancement. Students reported valuing AI for linguistic refinement, with one noting, “ChatGPT has helped me a lot with my English essays. When I am stuck or not sure about a word or idea, I ask ChatGPT.” Another described using AI for brainstorming: “Before I even start writing, I usually have a brainstorming session with ChatGPT.” A third student highlighted the confidence-building aspect of AI, stating, “knowing that I can rely on it to catch language errors or guide me when I am unsure has made essay writing less intimidating.” These findings illuminate why many language learners may become dependent on automated writing assessment tools. Similarly, Xiao and Zhi (2023) found that users highly valued ChatGPT’s accessible and adaptive feedback.

Studies in our review examined AI-generated feedback across several dimensions, including student perceptions, learning outcomes, and hybrid approaches combining teacher and AI feedback. Teacher feedback plays a crucial role in students’ writing development (Högemann et al., 2021), and students commonly prefer it over automated alternatives (Van der Kleij and Lipnevich, 2020). Nevertheless, AI was the principal source of feedback in most reviewed studies; only 11 involved instructor feedback. Among these, four examined educators in K-12 contexts and seven focused on higher education, with four studies exploring perspectives from both undergraduate students and instructors.

Ghafouri et al. (2024) investigated how ChatGPT-based writing instruction influences EFL teachers’ self-efficacy and learners’ writing skills. Results showed that teachers demonstrated significantly higher self-efficacy scores across measures of brainstorming, revising, providing feedback, and assessing, while students improved their writing skills on both post-test and delayed post-test measures. Ho (2024) interviewed ten students and four teachers in Hong Kong, finding hopeful yet skeptical perceptions: participants acknowledged AI’s helpfulness in educational feedback but expressed concern about its “generalized and ambiguous nature.” Only Guo and Wong (2023) directly compared feedback from ChatGPT and instructors. While ChatGPT generated more feedback in seconds, teachers and ChatGPT exhibited distinct styles in assessing specific properties of student writing. EFL teachers’ questionnaire responses ultimately revealed both positive and negative perceptions of using ChatGPT to support teacher feedback (Guo and Wong, 2023).

Two studies extended this inquiry to American Sign Language (ASL) learning. Hossain et al. (2021) qualitatively examined learner preferences for manual feedback from ASL experts vs. automatic feedback from ASLHelp, finding that students preferred human feedback in over half of cases. Conversely, Hassan et al. (2022) presented automatic feedback alongside videos of users practicing their signing, which received favorable ratings from participants. These contrasting findings suggest that while EFL learners consistently rate AI favorably, the effectiveness of AI feedback varies across different language learning contexts, particularly in visual-gestural languages like ASL.

AI feedback and peers

Four studies examined AI-assisted peer feedback in education. Lin et al. (2024) conducted a qualitative study involving undergraduate students in Canada to examine how “DD” (the chatbot) would guide peer reviewers to give more fruitful written feedback. After students interacted with the chatbot in a 50-minute session, researchers analyzed their interaction data, such as button clicks and text inputs. They focused on identifying eight possible interaction patterns and used these patterns to propose a methodology for refining chatbot scripts. They found that the chatbot effectively guided students in providing more structured and meaningful feedback. Their findings indicated that providing and receiving feedback led to similar levels of improvement in students’ revised essays; that is, their post-intervention scores showed improvement (Lin et al., 2024). Furthermore, this study introduces a methodology for refining chatbot scripts based on interaction pattern analysis, which can inform future designs of educational chatbots (Lin et al., 2024). In practice, educators may use chatbots like DD to scaffold tasks like peer review, where personalized instructor feedback may be limited.

Similarly, research involving 203 middle school students in the US examined the effectiveness of AI feedback by experimentally evaluating two approaches: combining supervised machine learning with natural language processing and ChatGPT, and ChatGPT evaluating peer feedback and providing feedback on feedback (Hutt et al., 2024). Hutt et al. found that ChatGPT’s feedback was less accurate such that there were considerable discrepancies between iterations in the wording of the responses provided. However, external measures strongly influenced both sources, e.g., students with higher metacognitive scores or those who valued mathematics more likely to give better feedback to their peers. Banihashem et al. (2024) coded feedback provided by peers vs. ChatGPT and found no significant relationship between the quality of essay writing and the feedback generated by the different sources. Tang et al. (2024) focused on AI influencing students' critical thinking regarding peer feedback, and to do so, they utilized the California Critical Thinking Disposition Inventory (Facione and Facione, 1992). The results show that while ChatGPT shows promise in assessing higher-level critical thinking skills, it still faces challenges in evaluating more nuanced aspects of critical thinking (Tang et al., 2024).

Bespoke AI feedback tools

Twenty-two studies examined bespoke AI systems and assessments designed specifically for educational feedback purposes (see Table 2). Other AI platforms investigated included programs built on PigAI, Grammarly, and Python (Shen et al., 2023; Yang et al., 2023; Sanosi, 2022; Leite and Blanco, 2020). When studying these locally developed AI programs, researchers examined perceptions, timing, accuracy, quality, and adaptability across various educational contexts.

Harati et al. (2021) explored student perceptions of the Assessment and Learning in Knowledge Spaces (ALEKS) system they developed. Using pre-test/post-test scores on an Adaptive Self-Regulated Learning Questionnaire (ASRQ), they measured changes in self-regulated learning (SRL) skills. Students reported low engagement with the platform, attributing this to the system's complexity, which made navigation and usage challenging (Harati et al., 2021). Furthermore, overall SRL scores decreased between the pre-test (mean = 88.07) and post-test (mean = 83.22), indicating that the ALEKS system failed to enhance students' self-regulatory capabilities. This represented one of the few studies demonstrating that a bespoke AI system did not adequately support students' learning needs.

In contrast, Cai et al. (2020) implemented an AI chatbot for Taiwanese students in Japanese language courses that utilized drama performances as a pedagogical approach. Researchers modified the Digital Reality Theater program in the experimental condition to incorporate an AI chatbot, while the control condition used the standard Digital Reality Theater without AI enhancement. Following implementation of the AI chatbot, they observed significant improvements in Japanese language proficiency within the experimental group.

Beyond ChatGPT, several other AI systems warrant attention. Guo et al. (2024) studied Eva, an AI-supported chatbot designed to facilitate peer feedback. Their results demonstrated that Eva successfully enhanced learners' investment in the peer feedback process. Leite et al. (2020) investigated AI-generated feedback on student coding assignments through an experimental design in which the control group received human feedback and the experimental group received automatic feedback. The researchers compared students' test and homework scores, final grades, and the nature of feedback (human vs. automatic), concluding that human feedback demonstrated only marginal superiority over automatic feedback. This finding reinforces the potential value of hybrid models combining human and AI-generated feedback.

Teacher perspectives on designing bespoke AI systems to assist with feedback provision were notably scarce in our sample. Of the 22 studies examining purpose-built AI, only Kong and Yang (2024) specifically focused on primary school teachers. These teachers participated in a six-week professional development program covering AI concepts, including machine learning, and strategies for integrating AI into instructional design. Participants completed pre- and post-assessments measuring AI perception and understanding across multiple dimensions, including technological and pedagogical content knowledge. Results revealed statistically significant increases across all assessed domains: both AI concept understanding and content knowledge improved substantially. Teachers also reported enhanced confidence in using generative AI tools for differentiated instruction and promoting student engagement (Kong and Yang, 2024).

When Lu (2019) developed a hybrid AI system called "Juku," an Automated Writing Evaluation (AWE) tool to assist Chinese students in writing courses, both student participants and teachers were asked for their perspectives on the AI-generated feedback. Results indicated that while most students (65%) found Juku helpful, a majority (60%) preferred a combination of AWE and teacher feedback. Teachers appreciated the time-saving benefits of Juku but raised concerns about its limitations in evaluating text structure, content logic, and coherence (Lu, 2019). This study reinforces AI's advantages in feedback timing while underscoring that it should not entirely replace human instruction.

Regarding the widely used ChatGPT, only one study in our sample directly compared a bespoke AI program against it. Solopova et al. (2022) analyzed reflective essays from the German Reflective Corpus, comparing feedback from ChatGPT-3 with their purpose-built system, PapagAI. Their results demonstrated that PapagAI substantially outperformed GPT-3 in emotion detection (0.81 vs. 0.49 accuracy) and Gibbs cycle classification (0.98 vs. 0.87 accuracy), highlighting the potential of specialized hybrid AI systems like PapagAI to enhance reflective practices in educational settings while addressing challenges posed by general-purpose large language models (Solopova et al., 2022). In the context of Germany's teacher shortage, these findings suggest how such tools could improve self-guided learning experiences for pre-service teachers while mitigating workforce challenges caused by high attrition rates (Solopova et al., 2022).

Six studies examined feedback from bespoke AI systems in EFL contexts. Three of these studies found insufficient statistical evidence of improved English learning outcomes (Krange et al., 2023; Lai et al., 2023; Zheng et al., 2024). Two studies reported enhanced EFL performance with the assistance of purpose-built AI systems (Liu et al., 2023; Salvador-Cisneros et al., 2023). Interestingly, Chen et al. (2022) found no differences in students' cognitive load or English-speaking skills when using their bespoke AI system. However, they observed decreased English-speaking anxiety among students who used the AI. Learning anxiety was measured using a questionnaire adapted from Thompson and Lee (2013), covering dimensions such as class performance anxiety, lack of self-confidence, and fear of ambiguity. Notably, the sample consisted of primary school students aged 10–11 years. These findings suggest that integrating dynamic assessment into speech recognition systems may effectively reduce learning anxiety, positioning such tools as promising resources for EFL education.

Expressed concerns and ethical considerations

The reviewed studies highlighted concerns about AI-generated feedback. Some studies raised concerns about the accuracy and consistency of ChatGPT-generated feedback, especially in complex tasks (Balse et al., 2023; Guo and Wang, 2023; Holderried et al., 2024; Mosoti, 2024; Naz and Robertson, 2024; Steiss et al., 2024; Téllez et al., 2024; Tossell et al., 2024). Studies also noted the absence of human connection and personalized guidance as a drawback, signaling the need for combined approaches (ElSayary, 2023; Escalante et al., 2023; Jukiewicz, 2024; Koutcheme et al., 2024; Lee, 2023; Leite and Blanco, 2020; Li and Kim, 2024; Lu, 2019; McGuire et al., 2024; Naz and Robertson, 2024; Ortega-Ochoa et al., 2024; Sanosi, 2022; Steiss et al., 2024; Téllez et al., 2024; Tseng and Lin, 2024; Wan and Chen, 2024; Zheng et al., 2024). There were concerns about students becoming overly dependent on ChatGPT and potentially misusing it for plagiarism or academic dishonesty (Alshammri, 2024; Mosoti, 2024; Tossell et al., 2024; Tseng and Lin, 2024; Wale and Kassahun, 2024; Werdiningsih et al., 2024).

Researchers also identified ethical considerations (Lai et al., 2023; Mosoti, 2024; Park, 2023; Tossell et al., 2024). Studies emphasized the need for clear guidelines and strategies in addressing academic integrity issues in unsupervised feedback generation. The studies do not acknowledge the possibility of bias in AI-generated feedback, particularly in working with diverse and neurodivergent student populations. Sharing student data with AI platforms raised concerns about privacy and the ethical use of student information (Almogren et al., 2024; Dai and Liu, 2024; ElSayary, 2023; Koutcheme et al., 2024; Morales-Chan et al., 2024; Mosoti, 2024; Solopova et al., 2023; Tang et al., 2024; Wang, 2024). One study suggested personalizing AI to the students' needs, as both tutors for low achievers and collaborative partners for high achievers (Xu et al., 2024). While the potential for ChatGPT to replace human teachers was identified, most studies advocated for a hybrid approach where AI assists and complements teachers' roles, ensuring human oversight and judgment in the learning process (Almogren et al., 2024; Fidan and Gencel, 2022; Chen et al., 2022; Colliot et al., 2024; Lu et al., 2024; Allen and Mizumoto, 2024; Koutcheme et al., 2024; Lu, 2019; Ortega et al., 2024; Mosoti, 2024; Tossell et al., 2024; Guo and Wang, 2023; Wang, 2024; Steiss et al., 2024).

Discussion

The development of AI represents a significant educational opportunity and challenge, given its increasingly pivotal role in supporting both students and educators. This review synthesized 104 studies published between 2008 and 2024, identifying key patterns in the integration of AI-generated feedback across educational contexts. In this discussion, we critically evaluate the state of the field by systematically distinguishing between evidence derived from stakeholder perceptions and evidence of actual learning outcomes, examining methodological strengths and limitations, addressing ethical and policy implications, identifying critical gaps, and situating our findings within the broader landscape of recent syntheses.

Differentiating perceptions from learning outcomes

A central contribution of this review is the clear differentiation between studies examining stakeholder perceptions of AI feedback and those measuring actual impacts on learning outcomes. This distinction is critical because favorable perceptions do not necessarily translate into meaningful learning gains, yet the literature often conflates these two dimensions.

Perceptions of AI feedback

Twenty-four studies in our sample examined perceptions of AI feedback, with findings generally indicating positive attitudes among both students and educators. Students frequently reported that AI feedback was personalized, accessible, and useful for revision (Alshammri, 2024; Xiao and Zhi, 2023). Educators, while more cautious, demonstrated growing openness to AI integration, particularly when they understood its capabilities and limitations (ElSayary, 2023; Kong and Yang, 2024). Notably, several studies found that students could not reliably distinguish AI-generated feedback from teacher-generated feedback (Almogren et al., 2024; Wan and Chen, 2024), and instructors modified GPT-generated feedback only about 30% of the time, suggesting a reasonably high degree of initial agreement with AI outputs.

However, perceptions alone provide an incomplete picture. As Van der Kleij and Lipnevich (2020) argued in the context of feedback research more broadly, perceptions must be examined alongside performance outcomes to draw meaningful conclusions about educational effectiveness. In our sample, only a subset of perception studies ($n = 9$) connected attitudinal data to performance measures, limiting the conclusions that can be drawn about whether positive perceptions reflect genuine pedagogical value or simply novelty effects.

Learning outcomes

Forty-nine experimental studies examined impacts of AI feedback on learning outcomes, with 23 employing controlled experimental designs and 25 incorporating pre- and post-test measures. The findings present a more complex picture than the perception literature alone would suggest. Several studies demonstrated positive effects of AI feedback on student performance, including improvements in reflective thinking, motivation, and domain-specific skills (Chiu et al., 2024; Wang et al., 2024b; Xu et al., 2024). However, multiple studies also found that teacher feedback remained superior to AI feedback (Steiss et al., 2024), and hybrid approaches combining AI with teacher input consistently outperformed AI-only conditions (Pahi et al., 2024).

Critically, the conditions under which AI feedback produces positive learning outcomes appear highly contingent. Studies that incorporated training for students or teachers, that involved teachers in the development or curation of AI feedback, and that employed paired approaches (AI plus human oversight) consistently yielded more favorable results than those implementing AI feedback as a standalone intervention. This pattern suggests that AI feedback is not a simple substitute for human feedback but rather a tool whose effectiveness depends on thoughtful integration into existing pedagogical structures. A recent systematic review mapping studies effects of AI-based feedback system supports this notion, calling for the development of hybrid human-AI feedback models (Durak and Onan, 2025).

Methodological strengths and weaknesses

The methodological diversity of the reviewed literature, encompassing experimental, quasi-experimental, mixed-methods, and qualitative designs, represents both a strength and a challenge for the field.

Strengths

A notable strength is the prevalence of rigorous experimental and quasi-experimental designs, which provide valuable causal evidence about the effects of AI feedback interventions. Forty-nine studies employed experimental approaches, many incorporating pre- and post-test measures and control conditions. Additionally, 87 studies collected data from multiple sources (e.g., combining surveys, interviews, and performance data), enhancing the triangulation of findings and providing richer insights into the mechanisms underlying observed effects.

Weaknesses

Despite these strengths, several methodological limitations warrant critical attention. First, sample size issues were pervasive. Several studies included very small samples, and in some cases, experimental research focused entirely on AI systems themselves with no human participants (Naz and Robertson, 2024; Solopova et al., 2023). Critically, none of the reviewed studies reported conducting power analyses to justify sample sizes, raising concerns about whether null findings reflect genuine effects or simply underpowered designs.

Second, the absence of longitudinal designs limits understanding of how AI feedback effects persist over time. The field is dominated by short-term interventions, with only a handful of studies including delayed post-test measures (e.g., Ghafouri et al., 2024). Given that feedback processes are inherently developmental, the lack of longitudinal data represents a significant gap.

Third, while many studies measured perceptions and performance, very few captured actual interaction data with AI systems. Only 13 studies collected data on how participants engaged with AI tools (e.g., clickstream data, interaction logs), leaving largely unexplored the behavioral mechanisms through which AI feedback exerts its effects. Future research employing process data, such as eye-tracking, keystroke logging, or behavioral trace data, could inform how students cognitively engage with AI-generated feedback during revision.

Fourth, despite the proliferation of experimental designs, the field lacks standardized outcome measures, making cross-study comparison and synthesis challenging. Heterogeneity in how learning outcomes, feedback quality, and perceptions are operationalized and measured complicates efforts to build cumulative knowledge.

Ethical considerations and policy implications

Our review identified significant ethical concerns that have received uneven attention in the literature and remain inadequately addressed by current policy frameworks.

Data privacy and security

A consistent concern across studies was the privacy and security implications of sharing student data with AI platforms. Several authors noted that when students submit work to AI tools, their intellectual property and personal data may be stored, analyzed, or used for model training without adequate transparency or consent (Almogren et al., 2024; Dai and Liu, 2024; Koutcheme et al., 2024). Institutional policies governing AI use in education remain nascent, and many institutions lack clear guidance on which AI tools are permissible for academic work, under what conditions, and with what data protections.

Algorithmic bias and fairness

A striking gap in the literature is the near-absence of attention to algorithmic bias, a core challenge for AI-based feedback systems identified in a review of AI-based feedback systems (Durak and Onan, 2025). None of the reviewed studies explicitly examined whether AI feedback systems perform differentially across demographic groups, learner characteristics, or neurodivergent populations. Given well-documented biases in large language models and the potential for automated feedback to perpetuate or amplify existing educational inequities, this oversight is concerning. Future research must systematically evaluate whether AI feedback systems produce equitable outcomes across diverse student populations, including learners with disabilities, English language learners, and students from marginalized socioeconomic and cultural backgrounds.

Academic integrity and over-reliance

Multiple studies raised concerns about students becoming overly dependent on AI tools, potentially undermining the development of independent critical thinking and writing skills (Alshammri, 2024; Mosoti, 2024; Tseng and Lin, 2024). The line between appropriate AI assistance and academic dishonesty remains contested, and few studies examined how to scaffold students' critical engagement with AI feedback rather than passive acceptance.

Policy implications

The findings of this review carry several implications for educational policy. First, institutions should develop clear guidelines for AI use in coursework that distinguish between permissible assistance and academic misconduct, and that require transparency from students about when and how AI tools are used. Second, teacher professional development programs must be expanded to equip educators with the skills to critically evaluate AI feedback, to integrate AI tools pedagogically, and to maintain human oversight of automated systems. Third, procurement policies for AI tools should require vendors to disclose data use practices, demonstrate fairness across student populations, and enable institutional auditing of algorithmic decision-making.

Gaps and future directions

Our review reveals several critical gaps that should shape the future research agenda.

Understudied populations

Higher education dominates the literature, with 76 studies focused on university students. Primary and secondary school students remain significantly underrepresented ($n = 13$). While this gap may reflect limited access to AI tools among younger students, neglecting earlier developmental stages restricts understanding of how AI feedback influences foundational skill development and how students' relationships with AI tools evolve over time. Echoing Durak and Onan (2025), we encourage researchers to examine AI feedback interventions in K-12 contexts, particularly in writing development where formative feedback is most consequential.

Geographic imbalance

The concentration of research in middle- and high-income countries, particularly China ($n = 30$) and the United States ($n = 15$), stands in stark contrast to minimal representation from Africa ($n = 2$) and Latin America ($n = 4$). This geographic imbalance limits understanding of how AI feedback functions across diverse educational systems, cultural contexts, and resource settings. Given that AI tools may offer particular promise for resource-constrained contexts where teacher-student ratios are high, the lack of research from low-income countries represents a missed opportunity.

Beyond perceptions

As noted earlier, the field requires more studies that connect perceptions to performance outcomes and that examine learning mechanisms rather than simply documenting attitudes. We extend Van der Kleij and Lipnevich's (2020) call to move beyond feedback perceptions toward rigorous examination of how AI feedback influences learning processes and outcomes.

Longitudinal research

The dominance of short-term intervention studies leaves unanswered questions about how AI feedback integration unfolds over time, whether initial effects persist, and how students' feedback literacy develops alongside AI tool use. Future research should employ longitudinal designs, capture behavioral process data, and investigate the conditions under which AI feedback produces durable learning gains.

Comparisons to earlier syntheses

The present scoping review complements and extends recent meta-analytic syntheses by providing a broader characterization of the research landscape on AI-generated feedback. Whereas Kaliisa et al. (2026) focused narrowly on comparative effectiveness, pooling 41 studies to examine whether AI-generated feedback outperforms human feedback in terms of learning performance and perceptions, our review encompasses 104 studies and captures a wider range of research questions, methodologies, and outcomes. Notably, Kaliisa et al. found no statistically significant difference in learning performance between AI and human feedback [Hedge's $g = 0.25$, CI (−0.11; 0.60)], with similarly null findings for feedback perception [Hedge's $g = -0.20$, CI (−0.67; 0.27)]. Our review corroborates these findings at a descriptive level: we identified multiple studies in which teacher feedback remained superior to AI feedback (e.g., Steiss et al., 2024), alongside others in which AI feedback produced comparable or enhanced outcomes when combined with instructional support (e.g., Chiu et al., 2024; Pahi et al., 2024). However, our review also reveals that the comparability of AI and human feedback is heavily contingent on contextual factors, including task complexity, implementation design, and the presence of teacher mediation, which are difficult to capture in meta-analytic effect-size aggregation.

In contrast to the outcome-oriented synthesis by Ba et al. (2025), which examined how AI-assisted feedback influences learner perceptions, actions, and outcomes through a theory-informed lens (the Self-System Model of Motivational Development), our review prioritizes methodological and contextual mapping. Ba et al. analyzed 129 studies published between 2014 and 2023, identifying a sharp rise in AI feedback research after 2018 and demonstrating that AI tools flexibly cater to multiple feedback foci and complexity levels. Their synthesis offers actionable insights into how AI can effectively enhance targeted learning outcomes. The present review extends this work by providing granular detail on the distribution of research methodologies (e.g., 49 experimental studies, 44 mixed-methods designs), participant demographics (e.g., 93 studies focused on students, only 12 on educators), and geographic representation (e.g., 55 studies from Asia, only two from Africa). This methodological mapping reveals that while the field has produced a robust body of experimental and quasi-experimental research, significant gaps remain, particularly in research involving younger student populations, low-income countries, and neurodivergent or marginalized learners, gaps that a meta-analytic approach may obscure.

Our review similarly extends the synthesis conducted by Durak and Onan (2025) focused on outcomes, benefits, and challenges of AI-based feedback systems. In their systematic review of 91, the researchers concluded that AI-based feedback systems seemed to support improvements across learning outcomes including motivation, self-regulated learning, and student-teacher interactions. Our research reinforces their mapping of the research landscape has primarily localized in higher education contexts and is supported by a high proportion of (quasi)experimental studies. Our scoping review builds on this mapping by articulating key geographical findings and highlighting core variables across the literature—importantly, ethical concerns, applications in peer feedback, and making a key distinction between research focused on perceived vs. observed outcomes.

Taken together, these complementary syntheses suggest a convergence of conclusions despite their differing methodological orientations. Both Kaliisa et al. (2026) and the present review advocate for hybrid approaches that pair AI-generated feedback with human oversight, recognizing the scalability of AI while preserving the deep, empathetic, and contextual features of human feedback. Similarly, Ba et al.'s (2025) call for clearer reporting of underlying AI algorithms and greater integration of pedagogical and technological insights aligns with our observation that only 13 studies in our sample collected data on actual AI interactions, and that few studies systematically examined the mechanisms through which AI feedback exerts its effects. Durak and Onan's (2025) calls for investigations of algorithmic bias and hybrid human-AI feedback approaches are echoed and extended in our findings through in-depth review of these concepts across the literature. Where meta-analyses provide precise estimates of average effects, our scoping review contributes a complementary understanding: the effectiveness of AI-generated feedback is not a fixed property of the technology itself but is strongly mediated by implementation context, methodological rigor, and the populations under study. As the field continues to proliferate, a pace that rendered our own review a historical snapshot, these combined insights underscore the need for future research to move beyond perceptions toward performance-based outcomes, to diversify participant samples, and to adopt longitudinal designs capable of capturing how AI feedback integration evolves alongside rapidly advancing technologies.

Limitations

Although the present review followed the PRISMA guidelines for scoping reviews, several limitations should be acknowledged.

First, the decision to include only papers published in English may have introduced language bias. Future reviews could expand the linguistic scope to capture relevant studies published in other languages, potentially revealing different regional or cultural perspectives on automated feedback in educational settings.

Second, the search was constrained by the number of databases used and by access limitations. Although ASReview's active learning algorithm accelerates screening by prioritizing relevant studies, this approach carries an inherent omission risk: studies ranked as less relevant by the algorithm were not manually verified, and it is possible that relevant records were excluded without review. Additionally, four full texts could not be retrieved despite repeated attempts; while these represented a small proportion of the sample, their exclusion may have introduced gaps in the evidence base, as their abstracts suggested potential relevance. Together, these factors introduce a degree of selection bias and reduce comprehensiveness, as some potentially relevant studies may have been omitted due to algorithmic prioritization or access constraints rather than eligibility criteria. Subsequent reviews could incorporate a broader range of databases, apply more exhaustive manual verification procedures, and utilize interlibrary loan or institutional collaborations to improve retrieval rates.

Third, our selection criteria excluded papers that examined educational feedback in non-active-learning settings. Future research could explore whether the patterns observed in this review hold across pedagogical contexts, including more traditional or lecture-based learning environments, to better understand how instructional format influences feedback practices.

Finally, the field of AI-mediated feedback is evolving rapidly, and the present review reflects the state of the literature at the time of data collection. Given the pace of publication in this area, we encourage readers to interpret this review as a historical snapshot of the evidence base, recognizing that new studies continue to emerge. Future updates or living reviews will be essential to maintaining currency in this dynamic field.

Conclusion

This scoping review provides a comprehensive mapping of the empirical literature on AI-generated feedback in education up to the time of data collection. We emphasize that this review represents a historical snapshot of a field that is evolving at an unprecedented pace. Since completing our synthesis, the literature on AI feedback has continued to proliferate, and newer papers are currently being reviewed as part of ongoing efforts to maintain currency with this rapidly advancing body of research.

The evidence mapped here suggests a cautiously optimistic picture, tempered by important caveats. The reviewed literature indicates that AI feedback shows promise for supporting learning outcomes, particularly when implemented alongside human instruction, though the mechanisms and conditions underlying these associations remain incompletely understood. The available studies also reflect significant heterogeneity in methods, outcomes, and quality, which limits the extent to which patterns identified here can be generalized. The evidence is consistent with the view that implementation context plays an important role in shaping how AI feedback is received and used, though the nature and direction of these relationships will require more controlled investigation to characterize fully. Critical gaps remain in the literature regarding how AI feedback functions across diverse populations and contexts, and the ethical implications of widespread AI integration warrant continued and urgent attention.

Given the pace of technological development and the accelerating growth of this literature, we encourage readers to interpret this review as a foundation upon which future updates will build. As newer studies continue to emerge, the conclusions drawn here will inevitably be refined, extended, and in some cases revised. Maintaining pedagogical rigor and critical evaluation will remain essential to ensuring that AI feedback serves, rather than replaces, the fundamentally human endeavor of education.

Author contributions

AL: Project administration, Validation, Methodology, Formal analysis, Writing – review & editing, Investigation, Software, Writing – original draft, Resources, Conceptualization, Supervision. CT: Validation, Writing – review & editing, Methodology, Formal analysis, Software, Visualization, Writing – original draft, Data curation. CD: Writing – review & editing, Conceptualization, Writing – original draft, Supervision. EN: Data curation, Methodology, Writing – original draft, Writing – review & editing, Formal analysis. NR: Writing – review & editing, Writing – original draft. TH: Writing – original draft, Supervision, Writing – review & editing, Conceptualization. EC: Writing – review & editing, Formal analysis, Writing – original draft, Validation. JM: Writing – original draft, Writing – review & editing.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors AL, CD declared that they were an editorial board member of Frontiers at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. We used ASReview to help with article screening. It is described in the manuscript and pre-registration

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2026.1799346/full#supplementary-material

References

*AlGhamdi, R. (2024b). Exploring the impact of ChatGPT-generated feedback on technical writing skills of computing students: a blinded study. *Educ. Inf. Technol.* 29 (14), 18901–18926. doi: 10.1007/s10639-024-12594-2

Allen, T. J., and Mizumoto, A. (2024). ChatGPT over my friends: Japanese English-as-a-foreign-language Learners' preferences for editing and proofreading strategies. *RELC J.* 57 (1), 89–106. doi: 10.1177/00336882241262533

*Almegren, A., Mahdi, H. S., Hazaea, A. N., Ali, J. K., and Almegren, R. M. (2024). Evaluating the quality of AI feedback: a comparative study of AI and human essay grading. *Innov. Educ. Teach. Int.*, 1–16. doi: 10.1080/14703297.2024.2437122

Almogren, A. S., Al-Rahmi, W. M., and Dahri, N. A. (2024). Exploring factors influencing the acceptance of ChatGPT in higher education: a smart education perspective. *Heliyon* 10 (11), e31887. doi: 10.1016/j.heliyon.2024.e31887

*Alshammri, A. (2024). ChatGPT in English writing: experiences and perceptions of Saudi EFL students. *Revista Amazonia Investiga* 13 (78), 47–57. doi: 10.34069/ai/2024.78.06.4

ASReview LAB developers. (2024). ASReview LAB - A tool for AI-assisted systematic reviews (v1.6.3rc0). *Zenodo.* doi: 10.5281/zenodo.11185216

Ba, S., Yang, L., Yan, Z., Looi, C. K., and Gašević, D. (2025). Unraveling the mechanisms and effectiveness of AI-assisted feedback in education: a systematic literature review. *Comput. Educ. Open* 9, 100284. doi: 10.1016/j.caeo.2025.100284

*Balse, R., Valaboju, B., Singhal, S., Warriem, J. M., and Prasad, P. (2023). "Investigating the potential of GPT-3 in providing feedback for programming assessments," in *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education Iticse*. 292–298. doi: 10.1145/3587102.3588852

*Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., and Drachsler, H. (2024). Feedback sources in essay writing: peer-generated or AI-generated feedback? *Int. J. Educ. Technol. High. Educ.* 21: 23. doi: 10.1186/s41239-024-00455-4

*Cai, M., Wang, J., Chen, G., Wang, J., and Yang, S. (2020). "A digital reality theater with the mechanisms of real-time spoken language evaluation and interactive switching of scenario & virtual costumes: effects on motivation and learning performance," in *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)*. 295–299. doi: 10.1109/icalt49669.2020.00096

*Carter, E. C., Schönbrodt, F. D., Gervais, W. M., and Hilgard, J. (2017). Correcting for bias in psychology: A comparison of meta-analytic methods. doi: 10.31234/osf.io/9h3nu

*Chen, C.-H., Koong, C.-S., and Liao, C. (2022). Influences of integrating dynamic assessment into a speech recognition learning design to support Students' English speaking skills, learning anxiety and cognitive load. *Educ. Technol. Soc.* 25 (1), 1–14. Available online at: https://www.jstor.org/stable/48647026 (Accessed October 31, 2024).

*Chen, Z., Zhu, J., McGrane, J. A., Hopfenbeck, T. N., Wang, Y., and Ji, Y. (2026). More inspiration and attention: how generative AI tools impact graduate students' affective engagement in L2 source-based academic writing. *Comput. Educ.* 242, Article 105417.

*Cheng, C. H. (2024). Using AI-generative tools in tertiary education: reflections on their effectiveness in improving tertiary students' English writing abilities. *Online Learn.* 28 (3) 33–54, doi: 10.24059/olj.v28i3.4632

*Chiu, M., Hwang, G., Hsia, L., and Shyu, F. (2024). Artificial intelligence-supported art education: a deep learning-based system for promoting university students' artwork appreciation and painting outcomes. *Interact. Learn. Environ.* 32, 824–842. doi: 10.1080/10494820.2022.2100426

*Claesen, A., Gomes, S. L. B. T., Tuerlinckx, F., and Vanpaemel, W. (2019). Preregistration: Comparing Dream to Reality. doi: 10.31234/osf.io/d8wex

*Colliot, T., Krichen, O., Girard, N., Anquetil, É., and Jamet, É (2024). What makes tablet-based learning effective? A study of the role of real-time adaptive feedback. *Br. J. Educ. Technol.* 55 (5), 2278–2295. doi: 10.1111/bjet.13439

*Cronje, J. (2023). Exploring the role of ChatGPT as a peer coach for developing research proposals: feedback quality, prompts, and student reflection. *Elect. J. e-Learn.* 22 (2), 1–15. doi: 10.34190/ejel.21.5.3042

*Dai, K., and Liu, Q. (2024). Leveraging artificial intelligence (AI) in English as a foreign language (EFL) classes: challenges and opportunities in the spotlight. *Comput. Human. Behav.* 159, 108354. doi: 10.1016/j.chb.2024.108354

*Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y., Gašević, D., et al. (2023). "Can large language models provide feedback to students? A case study on ChatGPT," in *2023 IEEE International Conference on Advanced Learning Technologies ICALT*. 323–325. doi: 10.1109/icalt58122.2023.00100

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., and Siemens, G. (2023). Impact of AI assistance on student agency. *Comput. Educ.* 210, 104967. doi: 10.1016/j.compedu.2023.104967

*Dong, L. (2024). Brave new world" or not?: a mixed-methods study of the relationship between second language writing learners' perceptions of ChatGPT, behaviors of using ChatGPT, and writing proficiency. *Curr. Psychol.* 43 (21), 19481–19495. doi: 10.1007/s12144-024-05728-9

Durak, H., and Onan, A. (2025). A systematic review of AI-based feedback in educational settings. *J. Comput. Soc. Sci.* 8 (4), 96. doi: 10.1007/s42001-025-00428-1

*Duval, S., and Tweedie, R. (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56 (2), 455–463. doi: 10.1111/j.0006-341X.2000.00455.x

ElSayary, A. (2023). An investigation of teachers' perceptions of using ChatGPT as a supporting tool for teaching and learning in the digital era. *J. Comput. Assist. Learn.* 40 (3), 931–945. doi: 10.1111/jcal.12926

*Escalante, J., Pack, A., and Barrett, A. (2023). AI-generated feedback on writing: insights into efficacy and ENL student preference. *Int. J. Educ. Technol. High. Educ.* 20 (1):57. doi: 10.1186/s41239-023-00425-2

*Evmenova, A. S., Regan, K., Mergen, R., and Hrisseh, R. (2024). Improving writing feedback for struggling writers: generative AI to the rescue? *TechTrends* 68 (4), 790–802. doi: 10.1007/s11528-024-00965-y

Facione, P. A., and Facione, N. C. (1992). *The California Critical Thinking Disposition Inventory (CCTDI)*. California Academic Press.

*Fidan, M., and Gencel, N. (2022). Supporting the instructional videos with chatbot and peer feedback mechanisms in online learning: the effects on learning performance and intrinsic motivation. *J. Educ. Comput. Res.* 60 (7), 1716–1741. doi: 10.1177/07356331221077901

*Fokides, E., and Peristeraki, E. (2024). Comparing ChatGPT's correction and feedback comments with that of educators in the context of primary students' short essays written in English and Greek. *Educ. Inf. Technol.* 30. doi: 10.1007/s10639-024-12912-8

Forsström, S, Njå, M., Munthe, E., Houldsworth, L., and Álvarez-Galván, J. L. (2025). Key findings and integration strategies on the impact of digital technologies on students' learning: Results from a literature review. OECD Education Working Papers.

*Ghafouri, M., Hassaskhah, J., and Mahdavi-Zafarghandi, A. (2024). From virtual assistant to writing mentor: exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill. *Lang. Teach. Res.* 00. doi: 10.1177/13621688241239764

Grassini, S. (2023). Shaping the future of education: exploring the potential and consequences of AI and ChatGPT in educational settings. *Educ. Sci.* 13 (7), 692. doi: 10.3390/educsci13070692

*Guo, K., Pan, M., Li, Y., and Lai, C. (2024). Effects of an AI-supported approach to peer feedback on university EFL students' feedback quality and writing ability. *Internet High. Educ.* 63, 100962. doi: 10.1016/j.iheduc.2024.100962

*Guo, K., and Wang, D. (2023). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Educ. Inf. Technol.* 29 (7), 8435–8463. doi: 10.1007/s10639-023-12146-0

Guo, W., and Wei, J. (2019). Teacher feedback and Students' self-regulated learning in mathematics: a study of Chinese secondary students. *Asia-Pac. Educ. Res.* 28 (3), 265–275. doi: 10.1007/s40299-019-00434-8

*Han, J., Liu, G., Liu, X., Yang, Y., Quan, W., and Chen, Y. (2024). Continue using or gathering dust? A mixed method research on the factors influencing the continuous use intention for an AI-powered adaptive learning system for rural middle school students. *Heliyon* 10 (12), e33251. doi: 10.1016/j.heliyon.2024.e33251

*Harati, H., Sujo-Montes, L., Tu, C., Armfield, S., and Yen, C. (2021). Assessment and learning in knowledge spaces (ALEKS) adaptive system impact on students' perception and self-regulated learning skills. *Educ. Sci.* 11 (10), 603. doi: 10.3390/educsci11100603

*Hassan, S., Lee, S., Metaxas, D., Neidle, C., and Huenerfauth, M. (2022). "Understanding ASL Learners' preferences for a sign language recording and automatic feedback system to support self-study," in *Proceedings of the 24th International Acm Sigaccess Conference on Computers and Accessibility Assets 2022*. 1–5. doi: 10.1145/3517428.3550367

Hattie, J. (2023). *Visible Learning: The Sequel. A synthesis of Over 2,100 Meta-Analyses Relating to Achievement* (1st ed.). London, UK: Routledge.

Hattie, J., and Timperley, H. (2007). The power of feedback. *Rev. Educ. Res.* 77 (1), 81–112. doi: 10.3102/003465430298487

*He, Y. (2024). The metaphor of AI in writing in English: a reflection on EFL learners' motivation to write, enjoyment of writing, academic buoyancy, and academic success in writing. *Int. Rev. Res. Open Distrib. Learn.* 25 (3), 271–286. doi: 10.19173/irrodl.v25i3.7769

*Holderried, F., Stegemann-Philipps, C., Herrmann-Werner, A., Festl-Wietek, T., Holderried, M., Eickhoff, C., et al. (2024). A language model-powered simulated patient with automated feedback for history taking: prospective study (preprint). *JMIR Med. Educ.* 10, e59213. doi: 10.2196/59213

Hopfenbeck, T. N., Zhang, Z., Sun, S. Z., Robertson, P., and McGrane, J. A. (2023). Challenges and opportunities for classroom-based formative assessment and AI: a perspective article. *Front. Educ.* 8, 1270700. doi: 10.3389/feduc.2023.1270700

*Hossain, S., Kamzin, A., Amperayani, V. N. S. A., Paudyal, P., Banerjee, A., and Gupta, S. K. S. (2021). Engendering trust in automated feedback: a two step comparison of feedbacks in gesture based learning. *Lect. Notes Comp. Sci.* 12748, 190–202. doi: 10.1007/978-3-030-78292-4_16

Högemann, J., Cunha, J., Núñez, J. C., Vallejo, G., Rodríguez, C., and Rosário, P. (2021). Writing intervention with elementary students struggling with writing: examining approach profiles to the teacher feedback on writing quality and motivational variables. *Read. Writ.* 34 (7), 1681–1710. doi: 10.1007/s11145-021-10159-0

*Hsia, L., Hwang, G., and Hwang, J. (2023). AI-facilitated reflective practice in physical education: an auto-assessment and feedback approach. *Interact. Learn. Environ.* 32 (9), 5267–5286. doi: 10.1080/10494820.2023.2212712

*Hu, Y. (2021). Effects and acceptance of precision education in an AI-supported smart learning environment. *Educ. Inf. Technol.* 27 (2), 2013–2037. doi: 10.1007/s10639-021-10664-3

*Hutt, S., DePiro, A., Wang, J., Rhodes, S., Baker, R. S., Hieb, G., et al. (2024). "Feedback on feedback: comparing classic natural language processing and generative AI to evaluate peer feedback," in *Fourteenth International Conference on Learning Analytics & Knowledge LAK 2024*. 55–65. doi: 10.1145/3636555.3636850

*Jukiewicz, M. (2024). The future of grading programming assignments in education: the role of ChatGPT in automating the assessment and feedback process. *Think. Ski. Creat.* 52, 101522. doi: 10.1016/j.tsc.2024.101522

*Kaliisa, R., Misiejuk, K., López-Pernas, S., and Saqr, M. (2026). How does artificial intelligence compare to human feedback? A meta-analysis of performance, feedback perception, and learning dispositions. *Educ. Psychol. (Lond)* 46 (1), 80–111. doi: 10.1080/01443410.2025.2553639

*Kim, M. K., Kim, N. J., and Heidari, A. (2022). Learner experience in artificial intelligence-scaffolded argumentation. *Assess. Eval. High. Educ.* 47 (8), 1301–1316. doi: 10.1080/02602938.2022.2042792

*Kong, S., and Yang, Y. (2024). A human-centered learning and teaching framework using generative artificial intelligence for self-regulated learning development through domain knowledge learning in K–12 settings. *IEEE Trans. Learn. Technol.* 17, 1588–1599. doi: 10.1109/tlt.2024.3392830

*Koutcheme, C., Dainese, N., Sarsa, S., Hellas, A., Leinonen, J., and Denny, P. (2024). "Open source language models can provide feedback: evaluating LLMs' ability to help students using GPT-4-as-A-judge," in *Proceedings of the 2024 Conference Innovation and Technology in Computer Science Education ITICSE*, 1, 52–58. doi: 10.1145/3649217.3653612

*Krange, I., Segaran, M., Gamlem, S., Moltudal, S., and Engeness, I. (2023). A triple challenge: students' identification, interpretation, and use of individualized automated feedback in learning to write English as a foreign language. *Deleted Journal* 59, 37–61. doi: 10.55612/s-5002-059-001

*Kurt, G., and Kurt, Y. (2024). Enhancing L2 writing skills: chatGPT as an automated feedback tool. *Journal of Inform. Technol. Educ. Res.* 23, 024. doi: 10.28945/5370

*Lai, P., Chen, J., Man, V., and Chan, C. H. (2023). "A new frontier in AI-assisted English oral presentation assessment," in *2023 IEEE International Conference on Teaching Assessment and Learning for Engineering Tale*, 4, 1–8. doi: 10.1109/tale56641.2023.10398320

*Lee, A. V. Y. (2023). Supporting students' generation of feedback in large-scale online course with artificial intelligence-enabled evaluation. *Stud. Educ. Eval.* 77, 101250. doi: 10.1016/j.stueduc.2023.101250

*Leite, A., and Blanco, S. A. (2020). "Effects of human vs. Automatic feedback on Students' understanding of AI concepts and programming style," in *SIGCSE 2020: Proceedings of the 51st Acm Technical Symposium on Computer Science Education*. doi: 10.1145/3328778.3366921

*Li, L., and Kim, M. (2024). It is like a friend to me: critical usage of automated feedback systems by self-regulating English learners in higher education. *Australas. J. Educ. Technol.* 40 (1), 1–18. doi: 10.14742/ajet.8821

*Li, T. W., Hsu, S., Fowler, M., Zhang, Z., Zilles, C., and Karahalios, K. (2023). "Am I wrong, or is the autograder wrong? Effects of AI grading mistakes on learning," in *Proceedings of the 2023 Acm Conference on International Computing Education Research V.1 ICER 2023 V1*. 159–176. doi: 10.1145/3568813.3600124

*Liang, H., Hwang, G., Hsu, T., and Yeh, J. (2024). Effect of an AI-based chatbot on students' learning performance in alternate reality game-based museum learning. *Br. J. Educ. Technol.* 55 (5), 2315–2338. doi: 10.1111/bjet.13448

Liao, X., Zhang, X., Wang, Z., and Luo, H. (2024). Design and implementation of an AI-enabled visual report tool as formative assessment to promote learning achievement and self-regulated learning: an experimental study. *Br. J. Educ. Technol.* 55 (3), 1253–1276. doi: 10.1111/bjet.13424

*Lin, M. P., Chang, D. H., and Winne, P. H. (2024). A proposed methodology for investigating student-chatbot interaction patterns in giving peer feedback. *Educ. Technol. Res. Dev.* 73, 353–386. doi: 10.1007/s11423-024-10408-3

*Liu, C., Liu, S., Hwang, G., Tu, Y., Wang, Y., and Wang, N. (2023). Engaging EFL students' critical thinking tendency and in-depth reflection in technology-based writing contexts: a peer assessment-incorporated automatic evaluation approach. *Educ. Inf. Technol.* 28 (10), 13027–13052. doi: 10.1007/s10639-023-11697-6

*Long, H. S. (2024). Exploring the use of ChatGPT as a tool for written corrective feedback in an EFL classroom. *J. AsiaTEFL* 21 (2), 397–412. doi: 10.18823/asiatefl.2024.21.2.8.397

Lu, Q., Yao, Y., Xiao, L., Yuan, M., Wang, J., and Zhu, X. (2024). Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing? *Assess. Eval. High. Educ.* 49 (5), 616–633. doi: 10.1080/02602938.2024.2301722

*Lu, X. (2019). An empirical study on the artificial intelligence writing evaluation system in China CET. *Big. Data.* 7 (2), 121–129. doi: 10.1089/big.2018.0151

*Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study. *Smart Learn. Environ.* 11 (1):9. doi: 10.1186/s40561-024-00295-9

McGuire, A., Qureshi, W., and Saad, M. (2024). A constructivist model for leveraging GenAI tools for individualized, peer-simulated feedback on student writing. *Int. J. Technol. Educ. (IJTE)*, 7 (2), 326–352. doi: 10.46328/ijte.639

Mosoti, A. K. (2024). *International College Students' Perceptions of Using ChatGPT in Producing Academic Essays*. Youngstown State University.

*Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* 4, 1. doi: 10.1186/2046-4053-4-1

Morales-Chan, M., Amado-Salvatierra, H. R., Medina, J. A., Barchino, R., Hernández-Rizzardini, R., and Teixeira, A. M. (2024). Personalized feedback in massive open online courses: harnessing the power of LangChain and OpenAI API. *Electronics. (Basel)* 13 (10), 1960. doi: 10.3390/electronics13101960

Naz, I., and Robertson, R. (2024). Exploring the feasibility and efficacy of ChatGPT3 for personalized feedback in teaching. *The Electronic Journal of e-Learning* 22 (2), 98–111. doi: 10.34190/ejel.22.2.3345

OECD (2026). *Navigating an Evolving Digital World*. Paris: OECD. Available online at: https://www.oecd.org/content/dam/oecd/en/about/projects/edu/the-pisa-2029-media-and-artificial-intelligence-(mail)-assessment-will-shed-light-on-whether-young-students-have-had-opportunities-to-learn-and-to-engage-proactively-and-critically-in-a-world-where-production,-participation,-and-social-networking-are-increasingly-mediated-by-digital-and-ai-tools-/PISA%202029%20MAIL%20Assessment%20Framework%20First%20Draft.pdf (Accessed December 1, 2025).

Ofqual (2024). *Ofqual's Approach to Regulating the Use of Artificial Intelligence*. Coventry, UK: Office of Qualifications and Examinations Regulation. Available online at: https://www.gov.uk/government/publications/ofquals-approach-to-regulating-the-use-of-artificial-intelligence-in-the-qualifications-sector/ofquals-approach-to-regulating-the-use-of-artificial-intelligence-in-the-qualifications-sector (Accessed September 01, 2025).

Ortega-Ochoa, E., Pérez, J. Q., Arguedas, M., Daradoumis, T., and Puig, J. M. M. (2024). The effectiveness of empathic chatbot feedback for developing computer competencies, motivation, self-regulation, and metacognitive reasoning in online higher education. *Intern. Things* 25, 101101. doi: 10.1016/j.iot.2024.101101

*Ouyang, F., Wu, M., Zheng, L., Zhang, L., and Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *Int. J. Educ. Technol. High. Educ.* 20 (1):4. doi: 10.1186/s41239-022-00372-4

Ouzzani, M., Hammady, H., Fedorowicz, Z., and Elmagarmid, A. (2016). Rayyan — a web and mobile app for systematic reviews. *Syst. Rev*. 5, Article 210. doi: 10.1186/s13643-016-0384-4

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). PRISMA 2020 Explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *Br. Med. J.* 372, n160. doi: 10.1136/bmj.n160

Paglialunga, A., and Melogno, S. (2025). The effectiveness of artificial intelligence-based interventions for students with learning disabilities: a systematic review. *Brain. Sci.* 15 (8), 806. doi: 10.3390/brainsci15080806

*Pahi, K., Hawlader, S., Hicks, E., Zaman, A., and Phan, V. (2024). Enhancing active learning through collaboration between human teachers and generative AI. *Comput. Educ. Open* 6, 100183. doi: 10.1016/j.caeo.2024.100183

*Park, J. (2023). Medical students' patterns of using ChatGPT as a feedback tool and perceptions of ChatGPT in a leadership and communication course in Korea: a cross-sectional study. *J. Educ. Eval. Health. Prof.* 20, 29. doi: 10.3352/jeehp.2023.20.29

*Parker, J. L., Becker, K., and Carroca, C. (2023). ChatGPT for automated writing evaluation in scholarly writing instruction. *J. Nurs. Educ.* 62 (12), 721–727. doi: 10.3928/01484834-20231006-02

Pishtari, G., Sarmiento-Márquez, E., Rodríguez-Triana, M. J., Wagner, M., and Ley, T. (2024). Mirror mirror on the wall, what is missing in my pedagogical goals? The impact of an AI-driven feedback system on the quality of teacher-created learning designs. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 145–156).

*Qiu, L. (2024). Feedback mechanism of English translation teaching based on deep learning. *J. Comput. Methods Sci. Eng.* 24 (4–5), 3313–3330. doi: 10.3233/jcm-247585

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

*Rad, H. S., Alipour, R., and Jafarpour, A. (2023). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: a case of wordtune application. *Interact. Learn. Environ.* 32 (9), 5020–5040. doi: 10.1080/10494820.2023.2208170

*Salvador-Cisneros, K. A., Bolanos-Mendoza, C. E., Bonini, F. X. D., and De Armijos, K. a. Y. (2023). "Application of AI in automatic feedback and rubrics for improving oral presentations in English as foreign language learners," in *Proceedings of 2023 6th International Conference on Educational Technology Management*. 1–7. doi: 10.1145/3637907.3637946

*Sanosi, A. (2022). To err is human: comparing human and automated corrective feedback. *Inform. Technol. Learn. Tools* 90 (4), 149–161. doi: 10.33407/itlt.v90i4.4980

*Sasaki, M., Mizumoto, A., and Matsuda, P. K. (2025). Machine translation as a form of feedback on L2 writing. *Int. Rev. Appl. Linguist. Lang. Teach*. 63 (4), 2301–2326. doi: 10.1515/iral-2023-0223

*Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *Br. Med. J.* 350, g7647. doi: 10.1136/bmj.g7647

*Shen, C., Shi, P., Guo, J., Xu, S., and Tian, J. (2023). From process to product: writing engagement and performance of EFL learners under computer-generated feedback instruction. *Front. Psychol.* 14, doi: 10.3389/fpsyg.2023.1258286

Shi, H., and Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL* 36 (2), 187–209. doi: 10.1017/S0958344023000265

*Shi, Y., and Deng, B. (2023). Finding the sweet spot: exploring the optimal communication delay for AI feedback tools. *Inf. Process. Manag.* 61 (2), 103572. doi: 10.1016/j.ipm.2023.103572

*Simonsohn, U., Nelson, L. D., and Simmons, J. P. (2014). p-Curve and effect size: correcting for publication bias using only significant results. *Perspect. Psycholog. Sci.: J. Assoc. Psychol. Sci.* 9 (6), 666–681. doi: 10.1177/1745691614553988

Solopova, V., Rostom, E., Cremer, F., Gruszczynski, A., Witte, S., Zhang, C., et al. (2023). "Papagai: automated feedback for reflective essays," in *In German Conference on Artificial Intelligence (Künstliche Intelligenz)*, (Cham: Springer Nature Switzerland), 198–206.

*Stanley, T. D., and Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Res. Synth. Methods.* 5 (1), 60–78. doi: 10.1002/jrsm.1095

*Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., et al. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learn. Instr.* 91, 101894. doi: 10.1016/j.learninstruc.2024.101894

*Tam, A. C. F. (2024). Interacting with ChatGPT for internal feedback and factors affecting feedback quality. *Assess. Eval. High. Educ.* 50 (2), 1–17. doi: 10.1080/02602938.2024.2374485

*Tan, X., Cheng, G., and Ling, M. H. (2025). Artificial intelligence in teaching and teacher professional development: a systematic review. *Comput. Educ.: Artif. Intell.* 8, 100355. doi: 10.1016/j.caeai.2024.100355

*Tang, T., Sha, J., Zhao, Y., Wang, S., Wang, Z., and Shen, S. (2024). Unveiling the efficacy of ChatGPT in evaluating critical thinking skills through peer feedback analysis: leveraging existing classification criteria. *Think. Ski. Creat.* 53, 101607. doi: 10.1016/j.tsc.2024.101607

*Téllez, N. R., Villela, P. R., and Bautista, R. B. (2024). Evaluating ChatGPT-generated linear algebra formative assessments. *Int. J. Interact. Multimed. Artif. Intell.* 8 (5), 75. doi: 10.9781/ijimai.2024.02.004

Thompson, A. S., and Lee, J. (2013). Anxiety and EFL: does multilingualism matter? *Int. J. Biling. Educ. Biling.* 16 (6), 730–749. doi: 10.1080/13670050.2012.713322

*Tossell, C. C., Tenhundfeld, N. L., Momen, A., Cooley, K., and De Visser, E. J. (2024). Student perceptions of ChatGPT use in a college essay assignment: implications for learning, grading, and trust in artificial intelligence. *IEEE Trans. Learn. Technol.* 17, 1069–1081. doi: 10.1109/tlt.2024.3355015

Tsai, C., Lin, Y., and Brown, I. K. (2024). Impacts of ChatGPT-assisted writing for EFL English majors: feasibility and challenges. *Educ. Inf. Technol.* 29, 22427–22445. doi: 10.1007/s10639-024-12722-y

*Tseng, Y., and Lin, Y. (2024). Enhancing English as a foreign language (EFL) Learners' writing with ChatGPT: a university-level course design. *Elect. J. e-Learn.* 22 (2), 78–97. doi: 10.34190/ejel.21.5.3329

UNESCO (2025). *AI and the Future of Education: Disruptions*. Paris: Dilemmas and Directions, United Nations Educational, Scientific and Cultural Organization.

*Van Campenhout, R., Kimball, M., Clark, M., Dittel, J. S., Jerome, B., and Johnson, B. G. (2024). "An investigation of automatically generated feedback on student behavior and learning," in *Fourteenth International Conference on Learning Analytics & Knowledge LAK 2024*. 850–856. doi: 10.1145/3636555.3636901

*Van Der Kleij, F. M., and Lipnevich, A. A. (2021). Student perceptions of assessment feedback: a critical scoping review and call for research. *Educ. Assess. Eval. Account.* 33 (2), 345–373. doi: 10.1007/s11092-020-09331-x

*Villagrán, I., Hernández, R., Schuit, G., Neyem, A., Fuentes-Cimma, J., Miranda, C., et al. (2024). Implementing artificial intelligence in physiotherapy education: a case study on the use of large language models (LLM) to enhance feedback. *IEEE Trans. Learn. Technol.* 17, 2025–2036. doi: 10.1109/tlt.2024.3450210

*Villegas-Ch, W. E., Govea, J., Gutierrez, R., and Mera-Navarrete, A. (2024). Improving interaction and assessment in hybrid educational environments: an integrated approach in microsoft teams with the use of AI techniques. *IEEE. Access.* 12, 93723–93738. doi: 10.1109/access.2024.3424397

*Wale, B. D., and Kassahun, Y. F. (2024). The transformative power of AI writing technologies: enhancing EFL writing instruction through the integrative use of writerly and google docs. *Hum. Behav. Emerg. Technol.* 2024, 1–15. doi: 10.1155/2024/9221377

*Wan, T., and Chen, Z. (2024). Exploring generative AI assisted feedback writing for students' written responses to a physics conceptual question with prompt engineering and few-shot learning. *Phys. Rev. Phys. Educ. Res.* 20 (1), 13. doi: 10.1103/physrevphyseducres.20.010152

*Wang, D. (2024). Teacher-versus AI-generated (poe application) corrective feedback and language learners' writing anxiety, complexity, fluency, and accuracy. *Int. Rev. Res. Open Distrib. Learn.* 25 (3), 37–56. doi: 10.19173/irrodl.v25i3.7646

*Wang, L., Chen, X., Wang, C., Xu, L., Shadiev, R., and Li, Y. (2023a). ChatGPT's capabilities in providing feedback on undergraduate students' argumentation: a case study. *Think. Ski. Creat.* 51, 101440. doi: 10.1016/j.tsc.2023.101440

*Wang, M., Long, T., Li, N., Shi, Y., and Chen, Z. (2024a). The impact of different types of feedback on pre-service teachers' microteaching practice and perceptions. *Educ. Inf. Technol.* 30, 5427–5448. doi: 10.1007/s10639-024-13024-z

Wang, N., Wang, L., Ma, Z., and Wang, S. (2022a). From knowledge seeking to knowledge contribution: a social capital perspective on knowledge sharing behaviors in online Q&A communities. *Technol. Forecast. Soc. Change.* 182, 121864. doi: 10.1016/j.techfore.2022.121864

*Wang, W., Lin, C., Lee, H., Huang, Y., and Wu, T. (2024b). Integrating feedback mechanisms and ChatGPT for VR-based experiential learning: impacts on reflective thinking and AIoT physical hands-on tasks. *Interact. Learn. Environ.* 33 (2), 1–18. doi: 10.1080/10494820.2024.2375644

*Wang, Y., Luo, X., Liu, C., Tu, Y., and Wang, N. (2022b). An integrated automatic writing evaluation and SVVR approach to improve students' EFL writing performance. *Sustainability* 14 (18), 11586. doi: 10.3390/su141811586

*Wang, Z., Denny, P., Leinonen, J., and Luxton-Reilly, A. (2023b). "Leveraging large language models for analysis of student course feedback," in *Proceedings of the 16th Annual Acm India Compute Conference Compute 2023*. 76–79. doi: 10.1145/3627217.3627221

*Werdiningsih, I., Marzuki, N., and Rusdin, D. (2024). Balancing AI and authenticity: eFL students' experiences with ChatGPT in academic writing. *Cogent Arts and Humanities* 11 (1), 2392388. doi: 10.1080/23311983.2024.2392388

*Xiao, Q. (2024). "ChatGPT as an artificial intelligence (AI) writing assistant for EFL learners: an exploratory study of its effects on English writing proficiency," in *Proceedings of the 9th International Conference on Information and Education Innovations ICIEI 2024*. 51–56. doi: 10.1145/3664934.3664946

*Xiao, Y., and Zhi, Y. (2023). An exploratory study of EFL learners' use of ChaTGPT for language learning tasks: experience and perceptions. *Languages* 8 (3), 212. doi: 10.3390/languages8030212

*Xu, Y., Zhu, J., Wang, M., Qian, F., Yang, Y., and Zhang, J. (2024). The impact of a digital game-based AI chatbot on students' academic performance, higher-order thinking, and behavioral patterns in an information technology curriculum. *Appl. Sci.* 14 (15), 6418. doi: 10.3390/app14156418

*Yan, D., and Zhang, S. (2024). L2 writer engagement with automated written corrective feedback provided by ChatGPT: a mixed-method multiple case study. *Humanit. Soc. Sci. Commun.* 11 (1), 1086. doi: 10.1057/s41599-024-03543-y

Yang, H., Gao, C., and Shen, H. (2023). Learner interaction with, and response to, AI-programmed automated writing evaluation feedback in EFL writing: an exploratory study. *Educ. Inf. Technol.* 29 (4), 3837–3858. doi: 10.1007/s10639-023-11991-3

Yuyang, Y., Liu, H., and Chau, T. (2025). A systematic review of AI ethics in education. *J. Glob. Inform. Manag.* 33 (1). doi: 10.4018/JGIM.386381

*Zheng, L., Long, M., Chen, B., and Fan, Y. (2023). Promoting knowledge elaboration, socially shared regulation, and group performance in collaborative learning: an automated assessment and feedback approach based on knowledge graphs. *Int. J. Educ. Technol. High. Educ.* 20 (1), doi: 10.1186/s41239-023-00415-4

Zheng, C, Chen, X., Zhang, H., and Chai, C. S. (2024). Automated versus peer assessment: Effects of learners' English public speaking. *Language, Learning and Technology (LL&T)*. 28 (2), 210–228. doi: 10.64152/10125/73578

*Zou, B., Du, Y., Wang, Z., Chen, J., and Zhang, W. (2023). An investigation into artificial intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills. *Sage. Open.* 13 (3), doi: 10.1177/21582440231193818