What are teachers' views on the application of AI in providing corrections and feedback on student essays?

Emmanuel Fokides, Eirini Peristeraki
fokides@aegean.gr, eiriniperister@gmail.com
University of the Aegean, Department of Primary Education

Abstract

Feedback to students' work is an important, yet challenging aspect of the educational process. In this respect, artificial intelligence (AI) may prove to be an invaluable tool in the hands of educators. Research on AI-generated feedback for student work exists, yet few studies have investigated educators' views on AI-generated feedback. For that matter, 20 primary school teachers corrected and provided feedback on 20 short essays coming from students aged eight to eleven. ChatGPT also provided feedback on these essays. The teachers compared their feedback with that of ChatGPT and subsequently participated in interviews. The findings revealed that while the participating educators acknowledged ChatGPT's technical capabilities that could potentially enhance various educational processes, they emphasized that it falls short in replicating the nuanced, empathetic interactions that human educators provide. They also emphasized that technology should complement, rather than substitute, the human touch in education.

Keywords: artificial intelligence, ChatGPT, educators, feedback, primary school, short essays

Introduction

Feedback, that is information provided by an agent, such as teachers, peers, or parents, regarding one's academic performance or understanding of a subject matter (Hattie & Timperley, 2007), has been extensively studied in educational research. Its importance in boosting students' learning is well-established (e.g., Hattie & Timperley, 2007; Henderson et al., 2019; Ryan et al., 2023). On the other hand, in order to provide comprehensive feedback, several factors have to be considered. For example, the manner/style in which feedback is presented significantly impacts student outcomes. It is essential to deliver feedback with kindness and discretion to enhance the performance of students, especially those encountering learning challenges (McLaren et al., 2011). Feedback should include constructive criticism while maintaining a balanced student-teacher relationship (Wang et al., 2008). Research indicated that the way praise is given needs careful thought, as it could reduce students' readiness to recognize their mistakes (Maclellan, 2005). General praise often lacks depth and might be counterproductive. Feedback must be tailored to meet the specific needs of each student, acknowledging their unique characteristics, linguistic, and cultural backgrounds (Henderson et al., 2019; Osakwe et al., 2022). Timeliness also plays a critical role; immediate feedback is more impactful than delayed feedback, which students might ignore if it becomes irrelevant to the content they are currently studying (Carless et al., 2011). Moreover, feedback's effectiveness varies based on individual and situational factors (Narciss et al., 2014). Given the above, educators often find it challenging to provide meaningful feedback (Crosthwaite et al., 2020) and sometimes resort to generic comments that lack depth (Weaver, 2006). Moreover, teachers frequently encounter systemic obstacles like strict school policies and heavy workloads, which can hinder their ability to give valuable and meaningful feedback (Hong, 2021).

Χ. Καραγιαννίδης, Η. Καρασαββίδης, Β. Κόλλιας, Μ. Παπαστεργίου (επιμ.), Πρακτικά Εργασιών 8ου Πανελλήνιου Συνεδρίου «Ένταξη και Χρήση των ΤΠΕ στην Εκπαιδευτική Διαδικασία», Πανεπιστήμιο Θεσσαλίας, Βόλος, 27-29 Σεπτεμβρίου 2024

Advancements in artificial intelligence (AI) are revolutionizing educational methodologies by streamlining tasks such as lesson planning, exercise creation, student assessment, grading, and feedback. Scholars, recognizing this potential, call for comprehensive research into AI's integration at diverse academic levels (Fuchs, 2023). Although incorporating AI into education presents challenges, evidence suggests its promising utility for educators (Lacey & Smith, 2023). Indeed, AI's application within education can substantially improve pedagogical effectiveness, enable continuous student progress monitoring, and dynamically adjust instructional methods to fit unique educational environments (Zhang & Aslan, 2021). However, there is a noted discrepancy between the current advancements in AI and educators' understanding of these developments, which could hinder AI's full potential in educational contexts (Chounta et al., 2022).

While there is research on AI-generated feedback for student work, as will be presented in the coming section, little focus has been given to its application on creative assignments like essays written in the context of language learning courses. Additionally, comparative studies between AI-generated and human-generated feedback are scarce and even fewer explored educators' views on AI-generated feedback. This led to the implementation of a project with two phases: the first involved a comparative analysis of AI- and human-generated feedback on short essays written by primary school students. The study at hand presents the second phase, where primary school teachers' opinions on both their own and AI feedback were analyzed. Details for this phase are presented in the following sections.

Related work

The body of scholarly work exploring AI integration in educational settings is on the rise. While skepticism over its utility and pedagogical alignment exists, others advocated for a more structured deployment of AI to prove its effectiveness (Gong et al., 2020). Focusing on automated task correction and grading, AI's role was examined in several studies. For instance, Grammarly, an AI-enhanced writing tool, significantly improved student performance by decreasing spelling and grammar errors (Sanosi, 2022). Nevertheless, its effectiveness in error correction appears limited to specific types (Fitria, 2021). The e-rater engine showed a 90% accuracy in correction and a .76 correlation with human grading (Azmi et al., 2019). Juku, an automated evaluation system for English education in China, received praise from students and teachers despite occasional lapses in evaluating structure, coherence, and content (Lu, 2019). ChatGPT demonstrated considerable reliability as a corrective tool and in predicting scores based on linguistic data, although it was suggested to be used alongside human correction for best results (Mizumoto & Eguchi, 2023). While valued for aiding in labor-intensive tasks (Mohamed, 2024), concerns about ChatGPT's impact on assessment validity and information authenticity were raised (Kiryakova & Angelova, 2023), with additional criticisms regarding its feedback quality (Moura & Carvalho, 2024).

Automated text scoring systems contribute to both consistency and objectivity in assessment protocols (Hussein et al., 2019). They enhance the pedagogical process by addressing grammatical and syntactic inaccuracies (Link et al., 2022). However, comprehensive text evaluation also considers parameters like relevance and coherence, areas that require further AI enhancement (Ramesh & Sanampudi, 2022). Automated systems also face difficulties in assessing creative writing and original expression, struggling with linguistic diversity and varying text structures across languages. As a result, their correction capabilities may not consistently match those of meticulous human evaluators (Murphy, 2019; Wang et al., 2022).

In terms of feedback provision, Jia et al. (2022) demonstrated that the Insta-Reviewer platform could produce feedback comparable to educators. Yet, in a literature review (Cavalcanti et al., 2021), it was found that while the majority of studies showed that automated feedback enhanced student performance (65.07%), a substantial number (46.03%) indicated that it does not significantly reduce educator workload. Moreover, there was no evidence suggesting that AI-generated feedback surpassed teacher-generated feedback in effectiveness (82.53% of studies). Several challenges accompany automated feedback systems, such as text overcorrection, cognitive overload, and inadequate explanations (Barrot, 2023). Additionally, they often cannot match human feedback due to non-specific recommendations, occasional inaccuracies, and unvarying responses in situations requiring unique feedback were also reported (Jia et al., 2022). It was suggested that AI systems need to offer feedback that is tailored to individual personality and language skills (Conati et al., 2021) and adopt a less strict style to enhance self-motivation and self-correction, especially beneficial for students with strong motivation or limited language skills (Liang et al., 2023).

Comparative studies on automated and teacher feedback, though limited, indicated that while automated feedback was often more detailed, students may disregard it. Teacher feedback positively impacted students' psychological well-being, but automated feedback excelled in enhancing language skills by focusing on grammar and syntax (Han & Sari, 2022; Wang & Han, 2022). To maximize benefits, integrating teacher insights with AI feedback has been proposed, allowing teachers to use AI tools to refine their assessments (Di Placito & Mortensen, 2023). While there is research related to educators' views and attitudes toward AI tools, there is far less research focusing on their views about the feedback provided by AI tools. As far as their attitude toward the integration of AI tools into their teaching is concerned, it seems that educators are increasingly aware of and generally positive, while there is no correlation between teaching style and attitude toward AI (Ghimire et al., 2024). Besides their positive attitude, they are also highly motivated to introduce AI-related content at school (Polak et al., 2022), though they have a moderate level of awareness regarding AI, their AI-related skills are still low, and they do not frequently use it in teaching (Alm & Ohashi, 2024). The educators recognized both the benefits and drawbacks of AI. Concerns included skepticism about AI substituting human educators, potential impediments to student development, the necessity to value human expertise, effects on critical thinking, risk of inaccuracies in AI-generated content, potential for cheating, and excessive dependence on AI platforms. The positive aspects noted were AI's ability to boost learner engagement, provide personalized learning, facilitate self-study, offer immediate feedback, and assist in content creation (Jose & Jose, 2024). In another study (Ohashia & Almb, 2023), it was found that although ChatGPT's utilization was minimal, the majority of educators showed interest in incorporating it into their teaching methodologies, specifically, for the development of language learning materials and favored its use for individual study over assessment tasks. Concerning the implications of ChatGPT in education, educators displayed both apprehension and optimism. While several concurred on its potential to enhance the accessibility of language education and benefit self-directed learning, a larger number expressed worries regarding possible academic dishonesty and over-reliance. Furthermore, the results of several studies indicated that training and better support are required in order for educators to successfully integrate AI effectively into their practices and to critically assess its applications while adhering to ethical standards. (Alm & Ohashi, 2024; Barrett & Packimply, 2023; Jose & Jose, 2024; Ohashia & Almb 2023; Tritscher et al., 2023).

As for educators' views on automated correction and feedback, the studies that have discussed the matter dealt primarily with the use of Grammarly. For example, participants

valued Grammarly's handling of high-level errors, which freed up time to focus on other aspects of their students' work. They were pleased by its comprehensive feedback, including its precise underlining of errors and detailed linguistic explanations, which aligned with their instructional approach. Additionally, users appreciated Grammarly's ability to highlight errors among individual students and the class, facilitating targeted improvements in both student assignments and broader teaching strategies (Koltovskaia, 2023). In another study (Ayan & Erdemir, 2023), the results revealed that most participants responded favorably to automated feedback and Grammarly. On the other hand, they noted inefficiencies because of incorrect vocabulary recommendations, its tendency to highlight the same grammatical mistakes numerous times, and its failure to provide efficient feedback in terms of coherence/cohesion, indicating that human touch is still needed. Otaki (2023) investigated the perceptions of students and educators in higher education regarding feedback provided by ChatGPT and by human educators. The analysis revealed themes highlighting the importance of understanding the nature of AI and human feedback, addressing the emotional dimensions of feedback, recognizing potential risks and ethical concerns, and exploring the integration of AI-generated feedback with human feedback practices to enhance learning engagement and outcomes. Finally, in a comprehensive study across 48 countries it was found that while many educators expressed intent to use ChatGPT for developing teaching materials, they displayed hesitation towards employing it for automated feedback and assessment (Alm & Ohashi, 2024).

Method

What can be concluded from the above presentation of the literature, is that the debate surrounding the pros and cons of AI-generated corrections and feedback is still unresolved. Moreover, the literature examining the views of educators regarding the correction and feedback provided by AI systems is rather limited. In light of these considerations, a study was conducted to examine the educators' views regarding AI's correction and feedback, having as an objective to answer the following research questions:

- RQ1a-b. According to educators which are (a) the positive aspects of ChatGPT's corrections/feedback and (b) which are theirs?
- RQ2a-b. According to educators which are (a) the negative aspects of ChatGPT's corrections/feedback and (b) which are theirs?
- RQ3. According to educators how ChatGPT's correction/feedback is compared to theirs?
- RQ4. Do educators prefer their corrections/feedback or ChatGPT's?
- RQ5. Would educators use ChatGPT to correct the work of their students?

Concerning the above RQs, the following have to be noted. ChatGPT was selected for its popularity and advanced capabilities compared to other AI systems. As will be presented in the following sections, certain stages of the study involved the correction of students' work by both ChatGPT and the participating educators. For that matter, it was decided to focus on short essays sourced from primary school learners. This educational level was selected because it focuses on essential literacy and numeracy skills, establishing the foundation for future academic success. Thus, the corrective feedback in this stage is crucial. Moreover, language skill development (central to primary education) is typically assessed through essays, which are integral to language courses in Greece. These essays, often produced in the classroom under timed conditions, yield brief and spontaneous texts. The brevity and

authentic nature of these compositions make them ideal for assessing linguistic proficiency and providing precise and meaningful feedback. Hence, the study targeted this type of essay for analysis.

The research adopted a qualitative method to explore the research questions, offering a unique contribution by not merely documenting participants' opinions. Unlike prior studies, this study involved educators in evaluating and commenting on student work, followed by contrasting their input with ChatGPT's. This process, details of which will be discussed in subsequent sections, ensured that their perspectives were grounded in direct experience and evidence rather than assumptions about ChatGPT's capabilities.

Participants

An invitation to participate was issued through social media, addressed to primary school teachers, detailing its aims and methods. Twenty teachers, all with over a decade of experience (M = 13.25, SD = 4.10 and aged 35-53 (M = 44.35, SD = 5.18), predominantly females (12 out of 20), agreed to participate. Regarding their understanding of ChatGPT's capabilities, one participant was knowledgeable but abstained from using it, five were aware but had not used it, and the remaining 14 were both informed and active users. The university's ethics committee approved the project, and informed consent was obtained from all participants before the study commenced.

Materials and instruments

Teachers unaffiliated with the study supplied short essays from students aged eight to eleven. From this collection, 20 were randomly chosen, with an average length of 200 words each. These handwritten essays, because of the subsequent requirement for analysis by ChatGPT, were transcribed verbatim into a digital document. The most recent version of ChatGPT available at the time of the study (v.4 turbo), was utilized for the correction of the essays and feedback provision. A precise and detailed prompt was essential for ChatGPT to effectively review and provide feedback on the essays. After several tests with a subset of essays and subsequent validations, the prompt was finalized. Due to limitations related to the length of the manuscript, readers can find the prompt, comments on it, and the full set of the results (themes, codes, and example quotes), in the following data repository: https://osf.io/bj638/?view_only=f9ad2f67f2454e9089058ed8dfc2de42. To mirror this process, educators used 20 Google Forms (one for each essay). These forms included the same correction instructions provided to ChatGPT, along with the essay to be corrected and fields for noting errors, providing feedback, and assigning a grade. Participants' views were captured through a structured interview, featuring the same questions as the study's RQs.

Procedure and data processing

The participating educators were given access to Google Forms to correct the 20 short essays within a day, accounting for potential delays due to the necessity of performing these corrections in a digital format. Subsequently, individual documents with educator feedback were created (20 participants X 20 short essays = 400 in total). The authors corrected students' essays using ChatGPT, with its outputs being compiled, once again, into a document. Its responses were reviewed against the guidelines given to it, to ensure relevance and accuracy. During the final stage, the participants received both their and ChatGPT's corrections/feedback and were given time (up to an hour) to review them before being interviewed. The interviews were conducted on an individualized basis and were recorded.

The transcribed interviews were thematically analyzed with NVivo v.1.7 by two skilled coders, to minimize the influence of subjectivity and to boost the overall reliability and credibility of the data interpretation process. They underwent training across multiple sessions utilizing a representative subset of interviews until they reached a high degree of intercoder reliability (Cohen's κ = 0.84). The results in the following section represent the coders' assessments as they were recorded in a final meeting in which the results from both were discussed.

Results and discussion

In this section, the results of the thematic analysis are presented and discussed. Please note that as there were cases in which participants' responses aligned with multiple codes, the cumulative frequency of coded instances surpasses the base count of 20, which corresponds to the number of participants involved in the study.

RQ1a-b. According to educators which are (a) the positive aspects of ChatGPT's corrections/feedback and (b) which are theirs?

According to educators, the positive aspects of ChatGPT's correction are multidimensional (Table 1). They recognized the efficiency of the tool in detecting a broad range of errors while remaining impervious to human fatigue and oversight. AI's ability to accurately detect mistakes has been noted in past research (e.g., Azmi et al., 2019; Mizumoto & Eguchi, 2023; Koltovskaia, 2023). They considered the constructive feedback provided by ChatGPT not only comprehensive but also one that aims to uplift and direct students towards continual improvement. ChatGPT's clarity and encouragement underscored its role as a facilitator of learning. This may lead to enhanced student performance as noted by Cavalcanti et al. (2021). Moreover, the technical advantage it possesses, from having extensive access to grammar resources to maintaining consistent accuracy, alongside its cost-effectiveness, positioned ChatGPT as a valuable asset within the educational domain. Indeed, consistency and objectivity in assessment are considered two of the AI's advantages (Hussein et al., 2019).

Table 1. Themes and codes for RQ1a

Theme	Interpretation	Codes	*n*	%
Efficiency in correction (31.7%)	The educators emphasized ChatGPT's promptness and thoroughness in identifying errors, as well as its resilience against human vulnerabilities such as fatigue.	comprehensive error detection	8	19.5
		unaffected by human limitations	3	7.3
		speed of correction	2	4.9
Constructive feedback (34.1%)	The educators appreciated ChatGPT's ability to provide detailed, rich feedback, which focused not only on the negatives but also on encouraging the student's progress.	rich feedback	5	12.2
		positive reinforcement	5	12.2
		detailed correction	4	9.8
Student-focused communication (17.1%)	The emphasis on ChatGPT's content being understandable and motivational highlighted its efficacy in directly addressing the students' needs through clear and supportive communication.	understandability/clarity	5	12.2
		motivational aspect	2	4.9
Technical advantage	The access to diverse educational resources, precision in correction, and	accuracy and precision	4	9.8
		access to resources	1	2.4

(17.1%)	cost-effective nature of ChatGPT underscored its superiority as a technological tool.	variety of evaluative options	1	2.4
		cost-efficiency	1	2.4

According to educators, the positive aspects of their correction involved a combination of personal empathy, constructive feedback, technical precision, approachability, and reflective practice (Table 2). Educators recognized the importance of building a supportive and empathetic relationship with students through personalized and humane corrections. Indeed, this constitutes a good practice, as empathy, kindness, and discretion, can enhance the performance of students (McLaren et al., 2011). They also emphasized the necessity of providing precise, targeted feedback that helps students understand their errors and how to improve. By doing so, they highlighted one of the key elements of effective feedback, which is the need to maintain a balance between being approachable, using simple language, and maintaining a supportive yet professional demeanor, that contributes to the effectiveness of feedback (Wang et al., 2008). Furthermore, the educators valued the human element of correction, which includes the understanding of the rationale behind errors and the expressive intentions of students. In essence, they supported that their feedback took into consideration the importance of providing feedback tailored to the unique characteristics of their students, which is another key element of effective feedback (Henderson et al., 2019; Osakwe et al., 2022), contrasting it with the rigidity of AI-based corrections.

Table 2. Themes and codes for RQ1b

Theme	Interpretation	Codes	*n*	%
Constructive and supportive feedback (24.2%)	The educators valued the importance of reinforcing students' confidence while also guiding them to recognize their errors, thus, maintaining student motivation.	positive feedback	3	9.1
		empathy	3	9.1
		encouragement of improvement	2	6.1
Personalization and understanding (33.3%)	Recognizing students' individuality, educators tailor their feedback and show a deep comprehension of student errors, which contrasts them with AI correction methods.	personalized approach	5	15.2
		comprehension of student errors	3	9.1
		human vs. AI correction	3	9.1
Balanced and targeted correction (42.4%)	In seeking to provide clear and constructive feedback, educators strive for a balanced approach that is neither overly critical nor dismissive, focusing on key areas for student development.	targeted correction	8	24.2
		objective and emotional balance	3	9.1
		leniency	2	6.1
		brevity in correction	1	3.0

RQ2a-b. According to educators which are (a) the negative aspects of ChatGPT's corrections/feedback and (b) which are theirs?

The negative aspects of ChatGPT's correction centered around three core themes: the lack of personalization in feedback, communication and language barriers, and inefficiencies in the correction process (Table 3). According to educators, ChatGPT failed to provide individualized feedback due to its inherent limitations in grasping the context and personal attributes of students. AIs' difficulties in understanding the context and in assessing creative writing, due to limitations in understanding the linguistic diversity and varying text structures were highlighted in several studies (e.g., Murphy, 2019; Wang et al., 2022). As noted in the preceding paragraph, personalization is an essential feature of effective feedback

(Henderson et al., 2019; Osakwe et al., 2022). Yet, as other studies noted (Conati et al., 2021), this is one of the AIs' limitations. Additionally, the use of complex language and a focus on positive aspects hindered effective communication and created barriers to student understanding. While, in past research, AI's detailed/comprehensive feedback was considered a positive feature (Koltovskaia, 2023), in this study, the educators argued that such detailed feedback can become incomprehensible to young learners. The participating educators were also concerned about the efficacy of ChatGPT's corrections, citing issues such as cognitive overload due to excessive information, as was also indicated in Barrot's (2023) study. Moreover, the omission of key errors and a tendency for overly strict grading were noted, confirming the findings of previous studies (e.g., Fitria, 2021; Jia et al., 2022; Liang et al., 2023). These findings indicate a need for ChatGPT to evolve toward a more nuanced, student-centered approach in its correction methodology to better align with educators' expectations and support student learning effectively, justifying the concerns expressed in previous studies (e.g., Conati et al., 2021; Liang et al., 2023).

Table 3. Themes and codes for RQ2a

Theme	Interpretation	Codes	*n*	%
Impersonalization, lack of contextual awareness (21,7%)	These codes reflect educators' concerns about ChatGPT's inability to understand students' individual contexts and to provide personalized feedback.	lack of personal understanding	3	13.0
		absence of human contact/ emotional intelligence	2	8.6
Inappropriateness of feedback and communication style (26.1%)	This theme encapsulated issues related to the suitability of feedback for the targeted age group and the balance of criticism and praise in communication.	inappropriate feedback for age	2	8.7
		insufficient conciseness	2	8.7
		lack of rigor	1	4.3
		overemphasis on positives	1	4.3
Issues in content and error analysis (39.1%)	Educators highlighted several aspects where ChatGPT's analytical abilities fall short, either through overloading information or missing out on nuanced errors.	error detection limitations	5	21.6
		misrecognition of prose expressions/everyday speech	2	8.6
		cognitive overload	1	4.3
		adherence to rules over meaning	1	4.3
Strict grading approach (13.0%)	The harshness of ChatGPT's grading, suggests a misalignment with educators' expectations.	strict grading	3	13

Educators identified several negative aspects of their corrections (Table 4). The main themes included inadequate feedback, human limitations, and correction inconsistency. Inadequate feedback was characterized by a lack of detail, brevity, and overlooked errors, signifying that educators felt they could improve in delivering comprehensive guidance to students. In essence, the participants concurred that they sometimes resort to generic comments that lack depth (Weaver, 2006). Human limitations highlighted issues such as bias and the complexity of language used, which might obscure the learning objectives. Lastly, correction inconsistency encompassed varied standards of assessment, including unfair or lenient scoring and the varying ability to maintain objectivity across multiple evaluations. These are in contrast with the objectivity and consistency of AIs (Hussein et al., 2019). Overall,

the analysis underscored the educators' reflective awareness of their correction practices and the areas where they believe enhancements are needed.

Table 4. Themes and codes for RQ2b

Theme	Interpretation	Codes	*n*	%
Inadequate feedback (33.3%)	These codes reflect the educators' acknowledgment of insufficient depth and comprehensiveness in their feedback to students.	lack of detail in feedback	5	20.8
		brief feedback	2	8.3
		lack of comments on emotional aspects	1	4.2
Human limitations (37.5%)	This theme encapsulated the educators' awareness of their vulnerabilities to error, biases, and the possible negative impact of complex language on students' comprehension.	omission of mistakes	4	16.7
		difficulty in finding positives	1	4.2
		human error	1	4.2
		influence of other texts	1	4.2
		time-consuming process	1	4.2
		use of advanced vocabulary	1	4.2
Correction inconsistency (29.2%)	Educators recognized that their ability to maintain consistent standards/fair judgment varied due to various factors, including time constraints and personal biases.	inconsistency in correction	2	8.3
		strictness in grading	2	8.3
		unfair grading	2	8.3
		leniency due to age	1	4.2

RQ3. According to educators how ChatGPT's correction/feedback is compared to theirs?

The educators' answers to this RQ effectively summarized the pros and cons of both theirs and ChatGPT's feedback examined in the previous RQs. In short, they perceived a mixed performance from ChatGPT (Table 5). While there was a common ground in error detection, educators expressed concerns over ChatGPT's grading rigor, detailed but potentially overwhelming feedback, and occasional omission of comments related to structural elements. They underlined the importance of human touch, emotional intelligence, and an appreciation for content and effort in their evaluations, which they often find lacking in ChatGPT's approach. Educators seek to balance corrective feedback with encouragement, often opting for a more lenient grading style that takes into account the developmental stage of their young learners. Therefore, while ChatGPT serves as a tool with certain advantages in terms of detail and explicative feedback, it falls short of completely capturing the empathetic and student-centered approach advocated by educators in the primary education context. Similar pros and cons were identified in other studies (e.g., Ayan & Erdemir, 2023; Otaki, 2023). Interestingly enough, the data did not reveal any codes or themes associated with AI's role in reducing educators' workload or decreasing the time needed for providing feedback, despite these aspects being highlighted as significant benefits of AI by other researchers (e.g., Koltovskaia, 2023). This can be attributed to the fact that the participants did not have access to the process of essay correction by ChatGPT; thus, they were not aware of the time needed to correct the essays, and, consequently, they were not able to make relevant comments. The only instance in which a code related to time management emerged was in the last RQ but in the context of balancing technology efficiency and the effectiveness of correction methods.

Table 5. Themes and codes for RQ3

Theme	Interpretation	Codes	*n*	%
Assessment accuracy (42.9%)	Educators and ChatGPT exhibit a shared basis for identifying errors, but divergent grading strategies and	differential grading approach	7	20.0
		similar error detection	6	17.1

	overlooking certain structural elements by ChatGPT indicate differences in assessment.	ChatGPT's omission of structural elements	2	5.7
Feedback quality (34.3%)	This theme revolved around the qualities of the feedback provided, ranging from the thoroughness of ChatGPT's analysis to the humanized and empathetic nature of educators' responses.	ChatGPT's detailed feedback	4	11.4
		human touch and emotion	4	11.4
		ChatGPT's encouragement discrepancy	2	5.7
		human conciseness/clarity	2	5.7
Educational values (22.9%)	Educators emphasized the importance of understanding content, students' effort, and the balance between grading leniency and content expression, highlighting a nuanced approach to teaching and evaluation	human understanding of content and effort	3	8.6
		human leniency in evaluation	3	8.6
		human content emphasis vs. rule adherence	2	5.7

RQ4. Do educators prefer their corrections/feedback or ChatGPT's?

The analysis of the educators' responses in this RQ suggested that they have a strong preference for their correction over ChatGPT's (Table 6). This preference is primarily driven by their belief in the irreplaceable value of the human touch in education, including the emotional and personal connections fostered between teacher and students, and the educator's deep understanding of student's abilities. However, there is also a significant indication that educators recognized the value of technology and were open to incorporating AI tools as a supplementary aid to enhance the correction process. The overarching sentiment underscored the importance of an approach where the educator's expertise is central, but technology serves as a valuable asset. Past research suggested that educators have little to moderate knowledge related to the capabilities of ChatGPT and AIs in general and that they do not frequently use such tools (e.g., Alm & Ohashi, 2024; Chounta et al., 2022; Ghimire et al., 2024). Contrary to that, the participants in this study were well-informed and most were active users of ChatGPT. Therefore, their opinion when faced with the dilemma of which feedback they prefer, can be somehow considered as more well-grounded. The clear preference for their feedback over ChatGPT's contrasts the findings of past research in which educators expressed a positive attitude toward automated feedback/correction (e.g., Ayan & Erdemir, 2023; Koltovskaia, 2023; Ohashia & Almb, 2023) and aligns better with research in which the participants expressed several concerns about the quality of AI-generated feedback and hesitation to use it (e.g., Alm & Ohashi, 2024; Otaki, 2023).

Table 6. Themes and codes for RQ4

Theme	Interpretation	Codes	*n*	%
Importance of personalized feedback (58.6%)	Educators stressed the need to understand and connect with students personally, which is something an AI cannot do.	preference for educator correction	9	31.0
		personal understanding of student	3	10.3
		need for the human element	3	10.3
		human holistic insight	2	6.9
The role of AI in correction (20.7%)	Some educators saw value in AI for its analytical abilities/detailed feedback that can complement their work.	utility of AI correction	3	10.3
		AI's analytical strength	3	10.3
Educator and AI synergy (20.7%)	Some educators believe that the ideal scenario includes a synergy of	complementary approach	4	13.8
		educator's correctional rigor	2	6.9

	AI's analytical strengths and educators' insights.

RQ5. Would educators use ChatGPT to correct the work of their students?
This RQ can be considered as an extension of the previous one. The analysis indicated that while there is openness to embracing new technologies for efficiency purposes, there remains a strong emphasis on the human-centric correction approach, highlighting the importance of personal interaction and emotional intelligence that technology lacks (Table 7). Educators valued their ability to understand and respond to the individual needs of students. The above, provide further justification for their preference of their feedback over ChatGPT's. Then again, there is a willingness to consider ChatGPT as a tool for certain tasks under specific conditions, while still ensuring human oversight and intervention; technology cannot replace their critical role in evaluating student work. This finding is in line with past research, in which the need for human touch was emphasized (e.g., Ayan & Erdemir, 2023). In addition, educators were concerned about the balance between saving time and maintaining the quality and effectiveness of corrections. The potential for time-saving is acknowledged, but only in circumstances where digital integration is sufficient and does not compromise the correction process. In sum, educators would not broadly use ChatGPT to correct the work of their students without serious consideration.

Table 7. Themes and codes for RQ5

Theme	Interpretation	Codes	*n*	%
Human-centric correction (42.3%)	The need for a personal, empathetic, and nuanced approach to evaluation that technology cannot replicate.	personal approach	8	15.4
		reluctance to use	8	15.4
		human element	6	11.5
Strategic utilization of technology (44.2%)	Some educators are open to integrating technology into their correction process but emphasize the importance of using it strategically and with human oversight.	review/oversight	9	17.3
		technology acceptance	8	15.4
		conditional use	6	11.5
Efficiency vs. effectiveness (13.5%)	Educators are considering the trade-offs between technology efficiency and the effectiveness of their correction methods	time management	6	11.5
		sufficient digital integration	1	1.9

Summary of the findings

To summarize the results, the study yielded critical insights into the views of educators regarding the correction and feedback mechanisms offered by ChatGPT compared to human ones. Educators acknowledged several benefits of utilizing ChatGPT, notably its efficiency, breadth of error detection, and the provision of clear, constructive feedback. ChatGPT's strengths lie in its unwavering consistency and technical proficiency. In contrast, human corrections were praised for their personalized, empathetic approach and intricate understanding of individual student needs. The ability of educators to blend technical accuracy with personal interaction and encouraging language represents the essence of effective educational feedback. Critique of ChatGPT centered on its failure to customize feedback, which often resulted in communication barriers and could lead to cognitive overload for students. Educators expressed a need for ChatGPT to adopt a more student-centered strategy in its feedback methods. In terms of human corrections, the primary concerns highlighted were inconsistencies, potential biases, and occasional inadequacies in feedback, stressing an opportunity for educators to enhance the thoroughness and clarity of

their feedback. The participants noted mixed performances from both ChatGPT and human corrections. While ChatGPT offered detailed and rigorous feedback, it sometimes lacked the necessary personal touch and flexibility that human educators provide, especially in understanding students' emotional and developmental nuances. There was a clear preference for the human over ChatGPT corrections and feedback among educators, rooted deeply in the value placed on personal relationships and emotional intelligence in the educational process. However, educators also recognized the supplementary benefits of ChatGPT and other AI tools in enhancing teaching and feedback methods. Although receptive to integrating AI technologies for their potential efficiencies, educators emphasized that any technological adoption should not undermine the essential human-centric approach fundamental to education. In short, there is a willingness to utilize AI tools under specific conditions where they support but do not supplant the educator's judgment and personalized interaction.

Limitations and future work

There are limitations in the study that need to be acknowledged. Choosing ChatGPT introduces uncertainty for other AI systems' performance. A more detailed/refined prompt might have been beneficial; yet complex prompts (resulting in more extensive and analytical feedback) could lead to biased results in favor of ChatGPT. The number and size of student essays were limited, affecting the depth of feedback by both educators and ChatGPT. Additionally, having a small number of participants may have narrowed the range of views that were recorded. These limitations should guide future research. Expanding the educator sample could capture a wider array of views. Including a greater number of essays on varied topics would enhance textual diversity for feedback provision. Investigating more detailed prompts, together with equally detailed guidelines for educators, is also advisable. Expanding the scope of the topics covered by interviews can offer a better understanding of educators' views. Finally, examining the efficacy of other AI systems is essential.

Conclusions

In conclusion, the study's results indicated that while ChatGPT was appreciated for its technical capabilities and potential to streamline certain aspects of educational feedback, it is evident that the subtle, empathetic engagement that human educators provide is irreplaceable. The study underscored an overarching educational philosophy where technology serves to augment, not replace, the human elements of teaching and learning.

References

Alm, A., & Ohashi, L. (2024). A worldwide study on language educators' initial response to ChatGPT. *Technology in Language Teaching & Learning, 6*(1), 1-23. https://doi.org/10.29140/tltl.v6n1.1141

Ayan, A. D., & Erdemir, N. (2023) EFL teachers' perceptions of automated written corrective feedback and Grammarly. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi, 5*(3), 1183-1198. https://doi.org/10.38151/akef.2023.106

Azmi, A. M., Al-Jouie, M. F., & Hussain, M. (2019). AAEE-Automated evaluation of students' essays in Arabic language. *Information Processing & Management, 56*(5), 1736-1752. https://doi.org/10.1016/j.ipm.2019.05.008

Barrett, A., & Pack, A. (2023). Not quite eye to AI: student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education, 20*(1), 59. https://doi.org/10.1186/s41239-023-00427-0

Barrot, J. S. (2023). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning, 36*(4), 584-607. https://doi.org/10.1080/09588221.2021.1936071

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, 36*(4), 395-407. https://doi.org/10.1080/03075071003642449

Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y. S., Gašević, D., & Mello, R. F. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence, 2*, 100027. https://doi.org/10.1016/j.caeai.2021.100027

Chounta, I.-A., Bardone, E., Raudsep, A., & Pedaste, M. (2022). Exploring teachers' perceptions of artificial intelligence as a tool to support their practice in Estonian K-12 education. *International Journal of Artificial Intelligence in Education, 32*(3), 725-755. https://doi.org/10.1007/s40593-021-00243-5 ©

Conati, C., Barral, O., Putnam, V., & Rieger, L. (2021). Toward personalized XAI: A case study in intelligent tutoring systems. *Artificial Intelligence, 298*, 103503.

Crosthwaite, P., Storch, N., & Schweinberger, M. (2020). Less is more? The impact of written corrective feedback on corpus-assisted L2 error resolution. *Journal of Second Language Writing, 49*, 100729. https://doi.org/10.1016/j.jslw.2020.100729

Di Placito M.L. & Mortensen E. (2023). Applying AI efforts to student assessments: That is, alternative innovations! *The Interdisciplinary Journal of Student Success, 2*, 93-108.

Fitria, T. N. (2021). Grammarly as AI-powered English writing assistant: Students' alternative for writing English. *Metathesis: Journal of English Language, Literature, and Teaching, 5*(1), 65. https://doi.org/10.31002/metathesis.v5i1.3519

Fuchs, K. (2023). Exploring the opportunities and challenges of NLP models in higher education: Is Chat GPT a blessing or a curse? *Frontiers in Education, 8*, 1166682. https://doi.org/10.3389/feduc.2023.1166682

Ghimire, A., Prather, J., & Edwards, J. (2024). Generative AI in education: A study of educators' awareness, sentiments, and influencing factors. *arXiv:2403.15586*.

Gong, X., Tang, Y., Liu, X., Jing, S., Cui, W., Liang, J., & Wang, F.-Y. (2020). K-9 artificial intelligence education in Qingdao: Issues, challenges and suggestions. *Proceedings of the 2020 IEEE International Conference on Networking, Sensing and* Control *(ICNSC)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICNSC48988.2020.9238087

Han, T., & Sari, E. (2022). An investigation on the use of automated feedback in Turkish EFL students' writing classes. *Computer Assisted Language Learning*, 1-24. https://doi.org/10.1080/09588221.2022.2067179

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Henderson, M., Phillips, M., Ryan, T., Boud, D., Dawson, P., Molloy, E., & Mahoney, P. (2019). Conditions that enable effective feedback. *Higher Education Research & Development, 38*(7), 1401-1416. https://doi.org/10.1080/07294360.2019.1657807

Hong, W. C. H. (2021). Macao secondary school EFL teachers' perspectives on written corrective feedback: Rationales and constraints. *Journal of Educational Technology and Innovation, 1*(1), 1-21.

Hussein, M. A., Hassan, H., & Nassef, M. (2019). Automated language essay scoring systems: A literature review. *PeerJ Computer Science, 5*, e208. https://doi.org/10.7717/peerj-cs.208

Jia, Q., Young, M., Yunkai, X., Jialin, C., Chengyuan L., Rashid, P., & Gehringer, E. (2022). Insta-Reviewer: A data-driven approach for generating instant feedback on students' project reports. *Proceedings of the 15th International Conference on Educational Data Mining*, 1-12. International Educational Data Mining Society. https://doi.org/10.5281/ZENODO.6853099

Jose, J., & Jose, B. J. (2024). Educators' academic insights on artificial intelligence: Challenges and opportunities. *Electronic Journal of e-Learning, 22*(2), 59-77. https://doi.org/10.34190/ejel.21.5.3272

Koltovskaia, S. (2023). Postsecondary L2 writing teachers' use and perceptions of Grammarly as a complement to their feedback. *ReCALL, 35*(3), 290-304. https://doi.org/10.1017/S0958344022000179

Kiryakova, G., & Angelova, N. (2023). ChatGPT-A challenging tool for the university professors in their teaching practice. *Education Sciences*, *13*(10), 1056. https://doi.org/10.3390/educsci13101056

Lacey, M. M., & Smith, D. P. (2023). Teaching and assessment of the future today: Higher education and AI. *Microbiology Australia*, *44*(3), 124-126. https://doi.org/10.1071/MA23036

Liang, K.-H., Davidson, S., Yuan, X., Panditharatne, S., Chen, C.-Y., Shea, R., Pham, D., Tan, Y., Voss, E., & Fryer, L. (2023). ChatBack: Investigating methods of providing grammatical error feedback in a GUI-based language learning chatbot. *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, 83-99. https://doi.org/10.18653/v1/2023.bea-1.7

Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning, 35*(4), 605-634. https://doi.org/10.1080/09588221.2020.1743323

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

Maclellan, E. (2005). Academic achievement: The role of praise in motivating students. *Active Learning in Higher Education, 6*(3), 194-206. https://doi.org/10.1177/1469787405057750

McLaren, B. M., DeLeeuw, K. E., & Mayer, R. E. (2011). A politeness effect in learning with web-based intelligent tutors. *International Journal of Human-Computer Studies, 69*(1-2), 70-79. https://doi.org/10.1016/j.ijhcs.2010.09.001

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), 100050. https://doi.org/10.1016/j.rmal.2023.100050

Mohamed, A. M. (2024). Exploring the potential of an AI-based chatbot (ChatGPT) in enhancing English as a foreign language (EFL) teaching: Perceptions of EFL faculty members. *Education and Information Technologies*, *29*(3), 3195-3217. https://doi.org/10.1007/s10639-023-11917-z

Moura, A., & Carvalho, A. A. A. (2024). Teachers' perceptions of the use of artificial intelligence in the classroom. In O. Titrek, C. S. De Reis, & J. G. Puerta (Eds.), Proceedings of the *International Conference on Lifelong Education and Leadership for All (ICLEL 2023)* (Vol. 17, pp. 140-150). Atlantis Press International. https://doi.org/10.2991/978-94-6463-380-1_13

Murphy, R. (2019). *Artificial intelligence applications to support K-12 teachers and teaching: A review of promising applications, challenges, and risks*. RAND Corporation, 1-20. https://doi.org/10.7249/PE315

Ohashia, L., & Almb, A. (2023). ChatGPT and language learning: University educators' initial response. *Proceedings of EUROCALL 2023*, 31-36. https://doi.org/10.4995/EuroCALL2023.2023.16917

Osakwe, I., Chen, G., Whitelock-Wainwright, A., Gašević, D., Pinheiro Cavalcanti, A., & Ferreira Mello, R. (2022). Towards automated content analysis of educational feedback: A multi-language study. *Computers & Education: Artificial Intelligence, 3*, 100059. https://doi.org/10.1016/j.caeai.2022.100059

Otaki, B. (2023). *Feedback in the era of generative AI: An explorative focus group study of student and educator perspectives in higher education* [Master's Thesis]. University of Gothenburg.

Polak, S., Schiavo, G., & Zancanaro, M. (2022). Teachers' perspective on artificial intelligence education: An initial investigation. Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, 1-7. CHI. https://doi.org/10.1145/3491101.3519866

Ramesh, D., & Sanampudi, S. K. (2022). An automated essay scoring systems: A systematic literature review. *Artificial Intelligence Review, 55*(3), 2495-2527. https://doi.org/10.1007/s10462-021-10068-2

Ryan, T., Henderson, M., Ryan, K., & Kennedy, G. (2021). Designing learner-centred text-based feedback: A rapid review and qualitative synthesis. *Assessment & Evaluation in Higher Education, 46*(6), 894-912. https://doi.org/10.1080/02602938.2020.1828819

Sanosi, A. B. (2022). The impact of automated written corrective feedback on EFL learners' academic writing accuracy. *Journal of Teaching English for Specific and Academic Purposes, 301*.

Tritscher, R., Röck, J., & Schlögl, S. (2023). Educ-AI-ted–Investigating educators' perspectives concerning the use of AI in university teaching and learning. In *International Workshop on Learning Technology for Education Challenges* (pp. 241-254). Springer Nature. https://doi.org/10.1007/978-3-031-34754-2_20

Wang, N., Johnson, W. L., Mayer, R. E., Rizzo, P., Shaw, E., & Collins, H. (2008). The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies, 66*(2), 98-112. https://doi.org/10.1016/j.ijhcs.2007.09.003

Wang, X., Lee, Y., & Park, J. (2022). Automated evaluation for student argumentative writing: A survey. *arXiv:2205.04083*. https://doi.org/10.48550/ARXIV.2205.04083

Wang, Z., & Han, F. (2022). The effects of teacher feedback and automated feedback on cognitive and psychological aspects of foreign language writing: A mixed-methods research. *Frontiers in Psychology, 13*, 909802. https://doi.org/10.3389/fpsyg.2022.909802

Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education, 31*(3), 379-394.

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers & Education: Artificial Intelligence, 2*, 100025. https://doi.org/10.1016/j.caeai.2021.100025