Al-Obaydi and Pikhart
Language Testing in Asia (2025) 15:38
https://doi.org/10.1186/s40468-025-00375-8

Language Testing in Asia

RESEARCH Open Access

AI partner versus human partner: comparing AI-based peer assessment with human-generated peer assessment in examining writing skills

Liqaa Habeb Al-Obaydi[1] and Marcel Pikhart[2*]

*Correspondence:
Marcel.pikhart@uhk.cz

[1] University of Diyala, Baqubah, Iraq
[2] University of Hradec Králové, Hradec Králové, Czech Republic

Abstract

This paper delves into the critical role of feedback in students' peer assessment, highlighting its variability across different educational settings and its profound potential to enhance learning outcomes, particularly within traditional and AI learning environments. It explores the evolution of feedback influenced by technological advancements such as AI, focusing on their application to improve EFL college students' writing skills through peer assessment. Using a rigorous a qualitative-dominant mixed method approach by integrating both quantitative and qualitative data, the study contrasts the effectiveness of traditional peer assessment (group A) with AI-based assessment using ChatGPT (group B) among fourth-year students from two different contexts, namely the University of Diyala, Iraq, and the University of Hradec Kralove, Czech Republic. While group A shows consistent improvement in writing skills, group B demonstrates slightly lower scores but offers quicker, accurate, and more precise feedback. The results of the study reveal significant differences between group A, utilizing traditional peer assessment, and group B, employing ChatGPT for AI-based assessment. In both contexts, group A's students demonstrated consistent improvement in writing skills, with final scores ranging from 7 to 14 out of 15. Group B also showed improvement, albeit slightly less pronounced, with scores ranging from 6 to 12 out of 15. Teachers' evaluations indicated that while group A benefited from reciprocal learning processes (writing and assessment), and greater social and cognitive engagement, group B experienced more accurate, quicker, and comprehensive feedback during peer assessment process from AI, albeit with less emotional and cognitive engagement. Ultimately, both methods contributed positively to students' writing skills, highlighting the strengths and trade-offs between human and AI feedback mechanisms.

Keywords: Feedback, Peer assessment, Foreign language learning, Artificial intelligence in education

Introduction

In order for students to develop and learn, feedback is vital (Selvaraj & Azman, 2020). Though in practice its influence is very variable, feedback has the potential to have a significant impact on students' achievement of learning outcomes (Jansen et al., 2024). It helps students improve their performance by updating their existing knowledge and actively adjusting their corresponding behavior, so it is a crucial component of the learning process (Al-Obaydi et al., 2023a, 2023b; Boud & Molloy, 2013; Colliot et al., 2024). In addition, researchers (Banihashem et al., 2024; Carless & Boud, 2018; Jansen et al., 2024) assert that feedback is a fundamental skill in the workplace and is not only connected to studies but also to lifelong learning. The frequency, format, and timeliness of feedback obtained from educational assessments have gradually altered to reflect the use of digital learning environments with assessment capabilities made possible by technological improvements (Jurs & Špehte, 2021; Ortega-Ruiperez & Correa-Gorospe, 2024). In order to offset this, there has been a rise in interest in the use of artificial intelligence (AI) in higher education for teaching and learning in recent years (Banihashem et al., 2024; Khosravi et al., 2022; Markauskaite et al., 2022). AI-powered advanced technologies allow students to provide various forms of feedback that enhance their learning and performance (Kohnke et al., 2023). Specifically, there has been an increase wave in discussing the use of AI to provide automated feedback on written tasks, with the goal of helping students become better writers (Conijn et al., 2020). One way to evaluate students' performance is through feedback, which can also be used as a peer assessment tool. Students evaluate one other's work and provide constructive criticism as part of peer assessment process. Both the reviewer and the person being reviewed can gain from this procedure (Babik et al., 2024).

Peer assessment is acknowledged as a scalable and sustainable form of evaluation that supports higher-order thinking and gives students prompt, thorough feedback on their work (Darvishi et al., 2022). It is sometimes difficult to have peer assessment in traditional face-to-face settings, but it can be done in a variety of ways when peer assessment is implemented using AI applications (Banihashem & Macfadyen, 2021), for instance. One of the most effective assessment techniques for enhancing students' writing assignments and learning is online peer feedback (Latifi & Noroozi, 2021). In such settings, students are not limited by time or location and are able to submit their writing and anonymously and reciprocally critique the works of their classmates (Lin, 2018; Ortega-Ruiperez & Correa-Gorospe, 2024). In particular, this setting makes it possible to incorporate several kinds of instructional tools (including scripting and guiding) that might help students give their peers more trustworthy, accurate, and pertinent replies (AlFarsi et al., 2021; Khaneh et al., 2022). In addition, lessening the workload of instructors in relation to student grading can allow for more enrollment chances for teachers.

Compared to traditional peer assessment, AI-based peer assessment may have a distinct impact on students' engagement. Research indicates that AI can improve participation by generating interactive learning experiences and offering real-time, tailored feedback (Salman & Chaya, 2024). On the other hand, there are worries that AI feedback could come across as impersonal and therefore lower emotional engagement (Lin & Chen, 2024). The way the AI tools are used and the personal preferences of each student can have a different effect on their engagement.

For many years, researchers have looked into automated writing evaluation as a means of quickly assessing student writing and relieving teachers of the task of grading papers (Ortega-Ruiperez & Correa-Gorospe, 2024; Wilson et al., 2022). There are a number of reasons why online peer feedback is necessary and its used is essential in teaching environment. Firstly, class sizes in higher education have increased (Banihashem & Macfadyen, 2021; Shahali Zadeh et al., 2016; Shi, 2019). Secondly, teachers' heavy workloads make it difficult for them to give students useful one-on-one feedback on their writings performance (Latifi & Noroozi, 2021), though some studies found negative impact of online teaching on writing skills (Al-Obaydi et al., 2025). According to Valero Haro et al. (2019), peer feedback has been presented as an efficient teaching technique in such settings to assist students in enhancing their writing. What is lacking from the literature, nevertheless, is information regarding how bachelor's students do in peer assessments to help their peers' writing performance (Yu et al., 2019).

The present study differs from the previous studies in many points; the first of them is that it deals with students of English language in to different contexts, Iraq and Czech Republic. It is possible that the socio-cultural and socio-educational environments of Iraq and the Czech Republic will have an impact on how EFL learners use peer assessment technique whether face to face or online, which may then have an effect on how they develop their writing skills in English. These two distinct contexts, Iraq and the Czech Republic, have been examined putting in mind that these two counties cannot accurately represent their respective regions; the selection process was conducted with the aim of assembling a sample that is both geographically and culturally diverse, thereby augmenting the generalizability and applicability of the study findings particularly with the use of technological applications, as recommended by Tarhini et al. (2015). The second point is the use of rubrics for writing assessment in both groups and the use of direct observation along the experiment which is not used by other studies.

The importance of studying AI partners versus human partners in peer assessment is to explore the potential benefits of AI-based feedback compared to traditional peer feedback in evaluating writing skills. This research aims to determine if AI can provide more consistent, objective, and personalized feedback, thereby enhancing learning outcomes and efficiency. It also examines the scalability and resource allocation benefits of AI, while evaluating its impact on students' development and integration into educational practices. By comparing both methods, the study seeks to identify which approach better supports student learning and skill development. Based on that, this study attempts to bridge the gap in literature in relation to the use of online feedback in the AI application in peer assessment of college students. The majority of studies on the use of AI in the classroom have not examined how these digital tools may affect learning results, instead concentrating on gauging students' attitudes about them (Haßler et al., 2016). The researchers' motivation stems from their observation that limited research that have looked into how using AI affects students' learning have produced conflicting findings and have just compared AI applications, not taking into account the possibility that the tools being used and their feedback vary depending on the context. With a combination of AI and human support, the current study aimed to thoroughly evaluate the impact of AI use on EFL college students peer assessment technique.

The study's rationale is based on the notion that peer assessment is a widely accepted method of helping EFL students improve their writing abilities while encouraging critical engagement and self-regulation in the learning process (Topping, 2018). The rise of ChatGPT and other artificial intelligence AI-powered technologies has led researchers and educators to investigate their potential for automated academic writing feedback (Ranalli, 2022; Zhai, 2023). Nevertheless, current research mostly focuses on either the overall advantages of AI-generated feedback or students' opinions of its usefulness, despite the growing interest in AI-assisted evaluation (Yu & Wang, 2021). It is critical to comprehend the efficacy of AI technologies in comparison to conventional peer feedback, given their growing integration in education. The specific goals of this study are to ascertain if peer assessments generated by AI produce quantifiable gains in writing skills more successfully than feedback generated by humans and to pinpoint the variables affecting students' participation in both forms of peer assessment. The results will help refine peer assessment methods for writing teaching in EFL contexts and add to the current discussion on AI uses in education.

The purpose of this study is to fill in a number of important gaps in the body of knowledge regarding peer evaluation as a form of feedback for EFL college students. Of particular interest is the relative efficacy of peer evaluation feedback produced by AI and that produced by humans.

Firstly, while numerous studies have explored the benefits and drawbacks of peer assessment and the use of AI in providing feedback (Shi & Aryadoust, 2024), there is limited research on how these two elements interact, especially among EFL learners in diverse socio-cultural contexts such as Iraq and the Czech Republic. Most existing studies have concentrated on the general advantages of AI in education or the broader impacts of peer assessment without delving into the nuanced differences in feedback quality and student perception when AI and human feedback are directly compared (Shi & Aryadoust, 2024). This study distinguishes itself by examining the two environments of Iraq and the Czech Republic, acknowledging that these settings may significantly influence how EFL learners perceive and utilize peer assessment and feedback. It is crucial to make clear that the comparison between Iraq with the Czech Republic does not primarily focus on cultural differences. The study intends to present a more comprehensive view of the effects of AI in education across many contexts by selecting these two discrete locations, which will enhance the results and provide more broadly applicable insights (Tarhini et al., 2015). This method enables a thorough assessment of AI's contribution to improving or impeding learning outcomes, independent of cultural factors. This geographical and cultural diversity in the sample helps in understanding the broader implications of using AI for educational purposes beyond a single cultural context, which is a limitation in many previous studies (2024).

Secondly, unlike other research that often overlooks the specific tools and frameworks used in peer assessment, this study incorporates the use of detailed rubrics for writing assessment in both AI and human feedback scenarios. The inclusion of rubrics ensures a structured and standardized approach to evaluating writing, which not only aids in maintaining consistency in feedback but also provides a clear benchmark for assessing the effectiveness and quality of feedback from both AI and human sources (Andrade, 2005). This methodological rigor is intended to provide a more precise comparison and

deeper insights into how structured guidance can impact the quality of peer assessments and the subsequent learning outcomes.

Lastly, by focusing on the practical applications and outcomes of AI versus human feedback, the study aims to provide actionable insights into how these different types of feedback can be integrated to support EFL students' peer assessment and writing development. The research questions specifically target the effectiveness, quality, and impact of feedback, aiming to elucidate whether a synergistic approach combining AI-generated and human-generated feedback can offer superior educational benefits compared to either method alone. This approach seeks to move beyond theoretical discussions and provide practical recommendations for educators looking to implement effective feedback mechanisms in diverse EFL learning environments. Therefore, the following three research questions were stated:

A) The impact of peer assessment on writing abilities:

1. What potential impacts might peer assessment feedback from AI partners (ChatGPT) have on EFL college students' writing abilities when compared to human-generated peer assessment feedback?

B) Quality of peer assessment feedback:

2. What are the differences in the perceived quality and usefulness of feedback provided by AI partners versus human partners among EFL college students?

C) Impact on the students' engagement:

3. What is the impact of AI-generated peer assessment feedback versus human-generated peer assessment feedback on the students' engagement including behavioral, emotional, cognitive, and social engagement?

Literature review

Feedback

Feedback is an essential component of the learning process, particularly in the context of developing writing skills among EFL college students (Al-Obaydi et al., 2023a, 2023b; Selvaraj & Azman, 2020). It serves to update students' existing knowledge and guide their subsequent actions, thus enhancing their academic performance (Yüksel et al., 2021). The effectiveness of feedback hinges on its timeliness, precision, relevance, and clarity, ensuring that students can understand and apply it effectively (Hattie & Timperley, 2007). When used in conjunction with iterative learning activities, feedback stimulates students' learning process and improves their comprehension of course materials (Eraut, 2006). Students utilize the feedback they receive to revise their knowledge of the task and, consequently, their course of action (Carless, 2019). Feedback must be timely, precise, pertinent to the students' performance, and easy for them to understand and apply in order to support their learning (Bulut et al., 2020). Technology breakthroughs have created new avenues for learning and altered the nature of education, such as the shift from traditional paper-and-pencil evaluations to digital online tests; the timing, format, and frequency of feedback obtained from educational assessments have also gradually altered to suit the demands of students (Jurs & Špehte, 2021). In recent years, the integration of AI in educational environments has revolutionized feedback delivery.

AI-powered systems, such as ChatGPT, provide immediate, detailed feedback covering a wide range of linguistic aspects, which is particularly beneficial for iterative learning activities (Klimova et al., 2024; Malik et al., 2022). However, while AI feedback and the use online learning generally excel in consistency and technical precision, it often lacks the emotional nuance and motivational elements inherent in human feedback (Derakhshan & Shakki, 2024). Human instructors can offer empathetic and encouraging feedback, which is crucial for maintaining student motivation and confidence (Lin & Chen, 2024). Therefore, a hybrid approach that combines the technical accuracy of AI with the personalized touch of human feedback could potentially offer the most comprehensive support for a student in a bended learning environment (Al-Obaydi, 2021). While human feedback attends to individual requirements, emotional support, and contextual understanding, AI provides precise, instantaneous, and scalable input that addresses cognitive components (Paiva & Bittencourt, 2020).

According to Zhang et al. (2019), AI can facilitate feedback practices in the education sector by offering feedback that is either fully or partially automated and can take several forms, like written feedback. In this concern, The United Nations Educational, Scientific, and Cultural Organization (UNESCO, 2019) states that it is critical to educate academics and practitioners on the current and potential uses of artificial intelligence (AI) in feedback practices, given the growing use of AI and modern technology in education worldwide. Researchers are divided between either to be positive and support the process of online feedback or to criticize it. In this concern, studies have indicated that feedback quality, student writing (Huisman et al., 2018; Noroozi & Hatami, 2019), the support of formative practices of feedback (Boud, 2020), feedback literacy (Carless & Winstone, 2020), domain-specific knowledge gain (Valero Haro et al., 2019), students' engagement (Al-Obaydi et al., 2023a, 2023b), and students' attitudes (Noroozi & Mulder, 2017) can all be enhanced by peer feedback. On the other hand, the application of peer feedback has drawn several major complaints. According to a number of studies, students' responses to feedback might differ greatly (Moore & MacArthur, 2016; Wilson et al., 2022). Concerns with peer feedback quality are raised by, who point out that students' limited experience, knowledge, and language skills may affect how useful it is. Additionally, Andriessen (2006) notes that there may be psychological and emotional repercussions to both giving and receiving critical comments.

Many new studies dealt with online feedback but in varied situations and contexts. For instance, Steiss et al. (2023) who investigated if generative AI could offer students constructive criticism on their compositions. By rating the comments each secondary school student provided on essays, they were able to compare the quality of human and AI criticism. The findings demonstrated that in every category except criteria-based, human raters performed a superior job of giving students insightful comments. Another new study by Jansen et al. (2024) dealt with comparing LLM (large language models) with expert-generated feedback without knowing the source of the input, 89 student teachers assessed the value of the feedback for their students' argumentative writing. Participants evaluated LLM-generated comments for text revision as helpful in 59% of texts (as opposed to 88% for expert feedback). Participants favored giving students'

comments created by LLMs in 23% of cases. Banihashem et al. (2024) compared the caliber of input generated by ChatGPT with that of peer feedback when students were tasked with writing argumentative essays. Seventy-four graduate students from a Dutch university made up the participant pool. In order to gather data for the study, two stages were involved: first, students' writings were collected while they were writing essays on one of the assigned subjects; second, peer and ChatGPT-generated feedback were gathered by having peers participate in a feedback process and by using ChatGPT as a source of input. The quality of essays and feedback was assessed using two different coding schemes: one for essay analysis and the other for feedback analysis. The outcomes demonstrated a noteworthy distinction between peer and ChatGPT feedback. Peers offered input that included details on identifying the essay's difficulty, whereas ChatGPT offered more descriptive feedback that included information about the essay's writing style. A comprehensive analysis of the data points to a possible synergistic function for students and ChatGPT in getting feedback. On other study by Colliot et al. (2024), three groups were compared: the pen and paper group, the pen and tablet group without feedback, and the pen and tablet group with feedback. An artificial intelligence-based teaching system that automatically deciphered participants' pen strokes on the screen gave the feedback. Eighty-five French students created three geometric forms on paper or on a tablet, and then they completed a paper transfer challenge. Comparing the tablet group to the paper and pencil group, the results indicated that while the tablet group's use did not improve learning, it did seem to increase interest in the job. In terms of both learning and transfer, the students in the tablet with feedback group outperformed the other two groups by a substantial margin. The current study is different from previous research in a number of ways. The first is that it focuses on English language learners in two distinct contexts: Iraq and the Czech Republic. Iraq and the Czech Republic's sociocultural and socio educational contexts might influence how EFL students employ peer assessment methods, whether in-person or online, which could subsequently influence how they improve their English writing abilities. The current study stands out from the others since it adds to the body of knowledge regarding AI in education and specifically use it in peer assessment. It also offers helpful guidance on how to use thoughtful feedback techniques to improve the writing skills of EFL college students.

Peer assessment

Peer assessment is a procedure where students examine one another's work and offer helpful criticism, encouraging critical thinking, deeper involvement, and cooperative learning (Topping, 2009). Recently, peer assessment has been employed as a highly effective tactic to raise students' writing standards (Huisman et al., 2019). A wide range of learning tasks, including oral presentations (Nejad & Mahfoodh, 2019), essays (Banihashem et al., 2024), portfolios (Yang et al., 2016), projects (Lin, 2018), and student-generated content (Darvishi et al., 2021), can be assessed by peers in a variety of settings, including online or in-class, pairs or group work (Topping, 2009), and open or blinded learning (Shoham & Pitman, 2021).

Peer assessment has many advantages as to strengthen the understanding of the subject matter (Li et al., 2010; Ortega-Ruiperez & Correa-Gorospe, 2024), acquire experience giving constructive criticism (Lundstrom & Baker, 2009), and grow in their sense of accountability. They also want to strengthen their ability to make critical decisions (Gyamfi et al., 2021; Khosravi et al., 2021). Assessment recipients are afforded the chance to obtain prompt and personalized feedback (Jansen et al. (2024) from peers with varying viewpoints (Patchan & Schunn, 2015). Peer assessment has emerged as a powerful pedagogical tool for enhancing students' writing skills through active engagement and mutual support among peers specifically in face-to-face peer assessment in class. It encourages students to critically evaluate each other's work, fostering a deeper understanding of the subject matter and honing their analytical skills. Successful learning outcomes are not guaranteed when students participate in the process without the right kind of support this is because these peers are viewed as less authoritative and more receptive to a mutual exchange of opinions and discussions (Topping, 2009).

Although peer assessment systems have numerous advantages, there are also typical worries and objections to their use. It is reasonable to argue that students may not have the necessary skills, motivation, or knowledge to fairly assess their peers' work and offer insightful comments. Several of the primary problems center on this idea of having students act as experts in training (Patchan et al., 2018). Darvishi et al. (2021) have demonstrated the ineffectiveness of basic aggregation approaches like mean and median, which assume equal weight for all reviewers, when used in peer assessment systems where multiple reviewers review the same task. This raises concerns about how to accurately assign a final grade. Another critical aspect of peer assessment is its impact on student engagement and learning behaviors (Matcha et al., 2019). As a result, a lot of teachers and higher education institutions have been hesitant to incorporate them into their curricula (Liu & Carless, 2006) and employ them in evaluations where the teacher's participation is crucial (Derakhshan et al., 2024).

The use of AI in facilitating peer assessment can address several challenges associated with traditional face-to-face settings, such as logistical constraints and inconsistent feedback quality (Chen et al., 2020). For instance, students can now obtain instant, tailored feedback from computerized examinations thanks to digital score reporting, which helps students make the most informed decisions about their education (Bulut et al., 2020). Though electronic correction for multiple choice questions sometimes lower students' cognitive skills, but it benefits teacher in saving time and efforts (Al-Obaydi, Pikhart & Tawafak, 2023). AI and electronic applications can streamline the process by providing structured frameworks and rubrics, ensuring that feedback is aligned with specific educational standards and reducing subjective bias (Michel-Villarreal et al., 2023). This standardization helps students give and receive more consistent and actionable feedback (Tempelaar, 2020). Moreover, incorporating rubrics in AI systems educates students on how to interpret and apply feedback effectively, promoting self-assessment and independent learning (Phuong et al., 2023). Despite these advantages, the human element in peer assessment remains invaluable. Human peers bring contextual knowledge and cultural sensitivity that AI systems may lack, making their feedback more relatable and nuanced (Patchan et al., 2018).

Online peer assessment, facilitated by AI, allows students to participate in the assessment process without being limited by time or location (Carless, 2019). This flexibility can lead to higher levels of engagement, as students can submit their writing and critique their peers' work at their convenience. Peer assessment conducted online can greatly increase student engagement by encouraging participation, responsibility, and teamwork (Al-Obaydi et al., 2023a, 2023b). It improves academic results by encouraging students to evaluate their own and their classmates' work critically, which broadens their comprehension of the subject (Lu & Law, 2012). Analyzing how students revise their work in response to AI versus human feedback in peer assessment can provide insights into the practical utility of each type of feedback provided. Studies have shown that while AI-generated feedback is detailed and immediate, students might find human feedback more trustworthy and motivating due to its personalized nature (Latifi & Noroozi, 2021). Therefore, understanding student perceptions and reactions to different feedback sources used in peer assessment is crucial for optimizing peer assessment practices. Additionally, cultural and contextual factors play a significant role in how feedback is received and utilized by EFL students in different regions, such as Iraq and the Czech Republic. Tailoring peer assessment approaches to reflect these cultural nuances can enhance their effectiveness and relevance, ultimately leading to better writing performance and learning outcomes.

AI feedback, peer assessment, and language learning intersect in a powerful way to enhance educational outcomes. With enhanced accuracy, AI can discover areas for development in language learning tasks by providing quick, accurate, and objective feedback (Wei, 2023). Diverse feedback sources are advantageous to students when combined with peer evaluation; peer reviews contribute unique viewpoints and personal insights, while AI provides consistency and speed (Darvishi et al., 2021). Students' ability to think critically and evaluate themselves is enhanced by this combination. Regarding language learning, this method promotes a thorough comprehension of language usage, fluency, and cultural quirks, which enhances speaking, writing, and comprehension abilities. Peer assessment combined with AI feedback produces a strong support network that encourages more involvement and more efficient language learning.

Methodology

Research design

The study employs a qualitative-dominant mixed method approach, integrating both quantitative and qualitative data to comprehensively analyze the effectiveness, quality, and impact of AI versus human peer feedback on EFL college students' writing skills. A qualitative-dominant mixed method is used to describe a research methodology that mostly uses qualitative data, but also adds quantitative data to support and improve the conclusions (Creswell, J. W. 2018). Quantitative data of the experiment, and qualitative data of the observations, are combined in mixed methods research. The approach can be categorized as qualitative-dominant mixed method approach, in which the quantitative components (the rubric scores) serve to further structure and validate the results, while the qualitative analysis (interpreting and comprehending the scores in context) takes center stage (Timans et al., 2019). Cross-validation and result corroboration are made possible by this combination (Creswell & Plano Clark, 2017).

The study is exploratory in nature (pilot study). The exploratory study design methodology, which permits a small sample size, is employed when researchers seek to obtain a deeper understanding of a particular phenomenon or a broader viewpoint (Creswell, 2018), as with the present study where the researchers attempt to see the difference between the two groups of the students who will engage in AI and human peer assessment. This kind of research is adaptable and unrestricted, which enables researchers to collect preliminary information, spot trends, and develop a fundamental comprehension of the topic (Kumar, 2019). Small sample size *t* might be used as evidence to support the dependability and validity of more extensive research. Small-scale sample insights offer important context, directing more thorough studies and guaranteeing solid, knowledgeable larger-scale research initiatives (Andrade, C. 2019). It indicates that the results might have broader applicability if comparable trends and outcomes are seen in various settings. Replicating the study in various contexts and accounting for regional differences might further confirm that the findings are generally applicable.

The research involves two groups of fourth-year EFL students from two different universities. Group A consists of eight students from the University of Diyala, Iraq (4 students), and the University of Hradec Kralove, Czech Republic (4 students), who will engage in human peer assessment. Group B comprises eight students as well from the University of Diyala, Iraq (4 students), and the University of Hradec Kralove, Czech Republic (4 students), who will engage in AI (ChatGPT) for peer assessment. Both groups have similar demographic attributes to ensure comparability, and efforts were made to control for variables such as previous academic performance and technology usage skills.

Ethical considerations include obtaining informed consent from all participants and ensuring their privacy and confidentiality throughout the study. The research received ethical approval from the Ethics Committee of the University of Hradec Kralove. This methodology aims to provide a detailed and rigorous analysis of the impact of AI and human feedback on the writing skills of EFL students, offering insights into the effectiveness and potential of AI in educational settings.

Proposed research plan

Figure 1 of Appendix 1 shows the plan that clarifies the questions of the study. It shows the systematic presentation of all variables within in class environment.

Participants

The focus of this study is on students in the fourth grade for two reasons; since they have been studying English as a foreign language for four years, their writing skills are advanced enough to participate in this experiment, and it is crucial to teach them how to utilize AI applications because they will soon graduate and enter the workforce. So, the participants are chosen purposefully (purposive sample) based on the above conditions. To guarantee that each student had an equal chance of being included in the study and to prevent bias, volunteers were chosen at random from a broader group. The first group of respondents were drawn from the University of Diyala, English Department of

the College of Education. Eight students total, four in each group, make up the sample. The second group of participants was a randomly chosen group of eight Czech students who study information and communication technology at the Faculty of Informatics and Management of the University of Hradec Kralove. The same tools to collect data were used both in the Iraqi and Czech group. The tasks were identical and the results were clustered together and analyzed.

The researchers do everything possible to guarantee more equivalency between the two groups in order to regulate the variables that could influence the experiment's outcome. Furthermore, the researchers made every effort to limit the impact of extraneous factors that might affect the experiment such as the previous academic level of the participants and whether they have outside work out of college time. The researchers are both the teachers of the participants and applied the experiment with the students by themselves without any help.

Demographic information reveals that 50% of the Iraqi participants are female students and 50% of them are male. The ages of the contestants varied from 22 to 23 years old. All participants were guaranteed adequate training on computer and mobile applications by the researchers. The same parameters were applied to the group of the Czech students: 4 male and 4 female participants, aged 22–23 years.

Tools

The main tool of the study is using AI application, specifically ChatGPT, to be the peer assessment partner as opposed to human partner. The participants used ChatGPT-4 and they use their own accounts. The students are divided into two groups, four for each in the two countries. The first group (group A) used traditional peer assessment process, each participant made three assessments for his/her other three participants' performance. Group B, tried to use AI applications, ChatGPT-4, to make the peer assessment for them. The rubric for group A is set, prepared, and validated by the researchers. The researchers exposed the rubric and the evaluation sheet to a group of international specialists to gain face and content validity. Several raters evaluated the identical set of study procedures in this investigation. Every expert agreed that the study's plan was appropriate, with only minor adjustments that were considered. A comprehensive and impartial review procedure was guaranteed by their independent evaluations. Interestingly, all experts agreed that the study plan was sound methodologically and conceptually. Only little changes were proposed, and these were quickly considered and included into the finished design. The final version of the tools after revision is as follows in Appendixes 1 and 2. Training for group A participants included instruction on how to give constructive criticism, practice scoring and standard alignment during calibration sessions, and comprehension of the evaluation rubric through thorough explanations and examples. This training improved the caliber of feedback given to students by guaranteeing consistency and dependability in their assessments. The reliability coefficients were calculated (0.879), which was an acceptable result.

In group B, there is no specific rubric but the student asks ChatGPT to revise his/her work based on the same points of group A's rubric. The student asked ChatGPT if it is possible to revise his/her work, then enter the work and ask it to revise it based on the same criteria of group A which are the main idea/focus, organization and format, language use and style, originality, and creativity. At the end, the student could ask ChatGPT to rate his work out of 15 to gain the last mark as with group A. The following are the questions that used as a guide and entered by the students into ChatGPT:

1. Could you please revise my composition based on main idea/focus, organization and format, language use and style, originality, and creativity?
2. Enter the composition.
3. Please mention my faults as points.
4. Could you rate my work out of 15?

By asking these questions, the students could get a detailed scoring for his/her work. The questions employed guarantee that the feedback produced by ChatGPT constantly satisfies the evaluation standards specified in the rubric. Furthermore, the criteria of the rubric are closely related to the input that ChatGPT offers. This improves the validity of the feedback because it accurately gauges the expected outcome. The teacher tries to enter the questions first with the students to show them the process to avoid any obstacles. The titles of composition are the same for the two groups in order to compare them in the results as follows:

1. A mobile phone could both be positive and negative
2. Newspaper reading is a good habit
3. In our busy lives, we have to express gratitude to teachers
4. Education makes an individual a better user of technology

It is important to mention that group A used paper-based composition writing while group B printed it because they need to enter it to ChatGPT. This process took much time from group B so the two groups are asked to write and print it from home to save time and efforts. The students worked under the supervision of the researchers who made sure that the feedback gained is identical. The students work only in college time where their teachers are available, and also, they trained previously to use the right prompts to ensure the same replies for both groups.

Another tool, evaluation sheet, is designed for teachers to evaluate the performance of the participants and the accuracy of both AI partner versus human partner in peer assessment process (see Appendix 2).

Data collection and analysis

The procedure begins with the assignment of five composition topics to both groups. For the peer assessment process, group A students assess their peers' compositions using a rubric developed by the researchers, with each participant performing one assessment for his/her peer (see Fig. 2 of Appendix 1). S refers to the students.

In contrast, group B students used ChatGPT to review their compositions based on the same criteria used with group A (see Fig. 3 of Appendix 1), which include the main idea/focus, organization and format, language use and style, originality, and creativity. S is the abbreviation of students who participated in the experiment.

Data collection includes also evaluation sheet which is an observation qualitative checklist constructed and used by the researchers to evaluate the performance of the two groups. Qualitative data is gathered through the evaluation sheet to understand students' behaviors on the feedback quality and usefulness to gain deeper insights into their experiences with peer and AI feedback. Qualitative data from the evaluation sheet will be subjected to thematic analysis to identify common themes regarding the perceived quality, engagement, and usefulness of feedback. The present study's thematic analysis entailed methodically locating, examining, and summarizing patterns (themes) in the qualitative data collected from the two teachers. The first step in the process is becoming acquainted with the data, which is followed by creating preliminary codes. After that, these codes are arranged into possible themes, which are then examined and improved to make sure they appropriately represent the data, as recommended by Creswell and Plano Clark (2017). Every criteria has a name and definition that clearly conveys its main ideas. During the analysis, Braun and Clarke's (2017) thematic analysis approach was consulted. The interview transcripts, in particular, were reviewed initially and then again multiple times. To familiarize himself with the concepts, the researcher immersed himself in the data. In order to uncover significant insights from the data, the themes are finally presented in a logical narrative that incorporates the analysis into a thorough report. Each researcher analyzed the data from his/her context then mixed together. The analysis begins with group A then group B, and at the end, they merged together.

Results

The following procedures are used to analyze the quantitative data: calculating the means for each composition separately, calculating the sum of each student's scores across all categories, then average them to determine overall performance. Lastly, to see the difference in performance, scores from several students (S1, S2, S3, and S4) are compared.

The results of the two groups are divided then compared to see the differences between the two groups. The two samples, Iraq and Czech, dealt with as one sample so there is no comparison or differences between the two contexts. Students are denoted by the letter S, and the ultimate score is out of 15. C stands for the four compositions.

Result of group A

Group A used peer assessment technique for assessing their peers' composition in class based on the rubric mentioned in Appendix 2. Each student assessed his/her three peers. As a result, we gained four scores for group A in Iraq and other four scores in group A in Czech Republic. S refers to students and the final score is out of 15. C refers to the compositions which are four. The process of assessment with group A was done smoothly without any obstacles in all the four times. Each student read the composition of his or her peer and assessed it with final score. The time spent for each assessment was about

half an hour or a bit more. The assessment of the four compositions took a month, once a week.

Table 1 of Appendix 1 shows that the final degrees of the Iraqi participants varied from 7 to 14. The most important result is that students' level increased exponentially with the repetition of writing and assessment processes as demonstrated in C3 and C4 for each student. The first marks of C1 are 7, 9, 12, and 12 respectively while the last marks of C4 are 13, 13, 12, and 14 respectively. This means that there is a kind of development in students' level of writing due to the use of peer assessment in class.

Table 2 of Appendix 1 demonstrates that the Czech Republic participants' final scores ranged from 8 to 14. The most significant finding is that, as each student's writing and assessment processes were repeated, their level rose exponentially, as seen in C3 and C4 for each student. The first scores for C1 are 9, 10, 9, and 9, and the final scores for C4 are 11, 13, 12, and 13, respectively. This indicates that the writing skill of the students has clearly improved.

Result of group B

Using the same criteria which are the rubrics listed in Appendix 2, group B evaluated their own composition using a peer assessment procedure with ChatGPT. As a result, we therefore received four points for group B in Iraq and an additional four points for group B in the Czech Republic.

The assessment procedure with group B went off without a hitch each of the four times. After printing their piece in accordance with the instructions, each student received a final score. Every assessment takes 5 to 10 min to complete. Once a week for a month, the four compositions were evaluated.

Using ChatGPT as a partner for peer assessment process, Table 3 of Appendix 1 demonstrates that the Iraqi participants' final scores ranged from 6 to 12. The most significant finding is that, as each student's writing and assessment processes were repeated, their level rose exponentially as with group A but a bit lesser, as seen in C3 and C4 for each student. The first scores for C1 are 7, 8, 6, and 7, and the final scores for C4 are 11, 11, 10, and 12, respectively. This indicates that the writing skill of the students has also clearly improved.

Table 4 of Appendix 1, which shows the final scores of the participants from the Czech Republic using ChatGPT as a peer assessment partner, shows that the scores varied from 6 to 12. The most important discovery is that, as each student's writing and assessment procedures were repeated, their level increased exponentially, albeit somewhat less than that of group A, as each student's C3 and C4 demonstrate. For C1, the initial scores are 7, 7, 6, and 8, while for C4, the final scores are 11, 11, 12, and 12. This suggests that the students' writing abilities have also obviously improved.

The outcomes of groups A and B in the two contexts, Iraq and the Czech Republic, are shown in Table 5 of Appendix 1. Table 5 of Appendix 1 concludes two major findings. The first is that the two groups' written composition scores show a discernible, steady rise. Second, despite the two groups appearing to be developed, ChatGPT rectification is more accurate so the degree of scoring is slightly lower than in group A which may be due to the neutrality nature (no passion) of AI applications. Group A is superior in terms of peer evaluation technical skills in class, but group B is superior in terms of AI

technological skills due to skill enhancement. It is important to note that the students in the two settings, Iraq and the Czech Republic, achieved scores that were almost identical; thus, there are no differences mentioned.

Results of the teachers' evaluation sheets

After some daily observations of the researchers and registering the performance of the students in all the four composition writing processes, and after making some periodical discussions with them, the researchers used the evaluation sheet, which is just like a behavior checklists, to evaluate the performance of the two groups based on the direct observation they made. In fact, the most reauthorization of the class observation, Reinke et al. (2013) includes additional recommendations for the administration of functional behavioral assessments in the context of school and mandates the completion of a school observation as part of the initial and reevaluation process for students (Appendix 3).

The summarized results of the two groups are clarified in Appendix 4. According to teachers' repeated observations, traditional peer assessment in class is very effective for the students because the process is reciprocal, as mentioned by one of the teachers "students benefit greatly from traditional peer assessment in the classroom because it creates a mutually beneficial learning environment. Through assessing others, students gain critical thinking and self-assessment abilities in addition to receiving comments from their peers on their level in writing." The quality of feedback in the use of AI as a peer assessment partner wins over ordinary peer assessment in class for many reasons such as it is accurate, quick, neutral (no passion), encouraging (provide encouraging feedback even the performance is weak), autonomous (helps for personalized learning), and comprehensive. Another teacher stated that "there are a number of benefits to using AI as a peer assessment partner over more conventional method. AI offers prompt, accurate feedback while removing human biases to guarantee neutrality and consistency. AI usually programmed to provide positive feedback even in cases where performance is subpar, which will help students stay motivated and develop." Biases and a lack of emotional connection are two issues with AI feedback that highlight serious challenges that need careful consideration. AI feedback biases have the potential to reinforce current disparities and produce skewed assessments based on the data used to train the systems. This may result in unfair or erroneous evaluations of students' performance. Furthermore, because AI feedback lacks emotional engagement, it is unable to provide the kind of nuanced support and encouragement that human teachers offer a critical component of student motivation and emotional health. According to the teachers, learning outcomes and writing performance in group A are better enhanced than group B because there is a reciprocal process of learning: writing and revising. In addition, social, emotional, and cognitive engagement existed more among the participants of group A due to the communication between participants and the objection for the final marks sometimes.

Discussion

The present study employed a qualitative-dominant mixed method approach, integrating quantitative and qualitative data to analyze the effectiveness of AI versus human peer feedback on EFL college students' writing skills. To ensure validation, reliability, and thorough feedback, the authors supplemented the experiment's students score findings with teachers' evaluations as recommended by Hattie and Timperley (2007). By detecting any biases and disparities in student ratings, teacher assessments offer an impartial way to improve the uniformity and standards of the peer review procedure. This method supports peer assessment goals and provides a more thorough, accurate assessment of student performance. The methodology ensured rigorous comparability between two groups of fourth-year EFL students from universities in Iraq and the Czech Republic. Group A engaged in human peer assessment, while group B utilized AI (ChatGPT) for peer assessment. Ethical considerations, including informed consent and privacy, were rigorously addressed, with approval from relevant ethics committee (The Ethics Committee of the University of Hradec Kralove, 2/2024). The study's comparison of AI and human aid in peer assessment makes it interesting. It is regarded as an expansion or application of earlier study findings to other contexts (the Czech Republic and Iraq). The study expands on previous research that has examined the advantages and disadvantages of peer assessment (e.g., Alqarni & Alshakhi, 2021; deBoer et al., 2023).

Quantitative results of the study indicated that group A (who received human feedback), generally achieved slightly higher scores compared to group B, which relied on AI-peer assessment feedback which is in agreement with Jansen et al. (2024). This is due to the fact that AI input was praised for its accuracy, speed, neutrality (no passion), and efficiency (Wei, 2023). The students in group A may be influenced by the emotional relationships with their peers so the results will not be neutral and the feedback looks more effective. This suggests that traditional peer assessment may have a slight advantage in terms of exchanging learning through writing and assessment. However, qualitative insights from teachers' evaluation sheets revealed that AI feedback in peer assessment was perceived as more accurate, quick, and comprehensive which are in line with Huisman et al. (2018) and Noroozi and Hatami (2019). In contrast, human feedback was valued for its reciprocity, social interaction, and deeper engagement, which contribute significantly to learning outcomes. Consequently, good feedback points students in the direction of significant progress by being precise, actionable, and unambiguous. While AI-generated feedback improved uniformity and objectivity and offer structured insights, human-generated feedback frequently provides them with more emotional connection and a deeper contextual understanding. The influence that feedback has on learning outcomes ultimately depends on how much students interact with and use it.

In the present study, the quality and perception of feedback varied between the two groups. AI feedback was noted for its accuracy, speed, neutrality (no passion), and efficiency as with Wei (2023), but lacked the nuanced understanding and social interaction inherent in human feedback and in addition to the biases that may be occurred. Human feedback, on the other hand, offered richer exchanges that fostered emotional and cognitive engagement among students, though it sometimes lacks the enough knowledge

from the side of the students, as concluded by. This reciprocal interaction in group A appeared to enhance learning outcomes by encouraging iterative processes of writing, revision, and peer assessment.

In terms of learning engagement, group A participants demonstrated higher levels of social interaction, emotional involvement, and cognitive engagement during the peer assessment process. These factors were attributed to direct peer interaction and discussions (Ortega-Ruiperez & Correa-Gorospe, 2024). In contrast, group B participants showed higher behavioral engagement due to the novelty and technical appeal of using AI for assessment which is as supported by Tawafak et al. (2024) related to the process of technology integration between the students and the application used. Online applications generally could provide a sense of learning engagement specifically if there are cooperative learning environment (Al-Obaydi et al., 2023a, b).

This research adds to the increasing amount of knowledge on peer assessment when the partner uses AI or human support, as well as the application of analytical rubrics to enhance learning results. This study offers useful evidence-based insights that can direct the development and implementation of AI applications in EFL settings by assessing the efficacy of analytical rubrics for peer assessment when the partner uses AI or human assistance within the context of EFL teaching and learning in Iraq and the Czech Republic.

The practical implications of these findings suggest that while AI can provide efficient feedback, it should be integrated thoughtfully with human interaction in educational settings (Banihashem et al., 2024) in what is called blended learning environment that mix between human and technology in teaching (Al-Obaydi, 2021). This combination can leverage AI's strengths in efficiency and objectivity while harnessing human feedback to foster deeper learning and interpersonal skills and to overcome the risk and uncertainty that associated with the use of AI applications in teaching (Pikhart & Al-Obaydi, 2025). Moreover, integrating AI tools like ChatGPT can help students develop technological skills, but ongoing support and guidance by teachers are essential to ensure effective use and integration into educational practices (Tawafak et al., 2023, 2024). It is recommended to use both human and AI feedback to present a comprehensive blended learning environment. Quick, data-driven insights can be provided by AI, but if given precise instructions, human feedback can specifically address contextual and emotional elements, and finally, given the educators' professional development opportunities to acquaint them with AI resources and providing them with the tools they need to decipher and add their knowledge to AI responses (Ali et al., 2023).

Future research directions could include longitudinal studies to examine the sustained impact of AI and human feedback on writing skills and overall academic performance. Cross-cultural comparisons could also explore how cultural factors influence the effectiveness of AI in different educational contexts. Furthermore, future research must carefully analyze and resolve procedural bias resulting from differences in task execution, such as paper-based versus ChatGPT-based approaches, such as mode of interaction and feedback quality, in order to achieve a fair comparison. Furthermore, continuous improvement of AI algorithms to simulate more nuanced human-like feedback and interaction remains a critical area for advancement in educational technology.

In conclusion, while AI presents promising benefits in educational assessment, human peer feedback remains indispensable for nurturing holistic learning experiences and skill development among EFL students. Balancing these approaches strategically can optimize educational outcomes and prepare students effectively for the challenges of a technologically advancing world.

Conclusion

Feedback is essential for student development and learning, playing a critical role in achieving academic success. Although its impact can vary, effective feedback significantly influences students' performance. Beyond academics, feedback is a crucial skill in the workplace, contributing to lifelong learning. With advancements in technology, the nature of feedback has evolved, leading to changes in its frequency, format, and timeliness. This evolution has sparked interest in using artificial intelligence (AI) in higher education for teaching and learning. AI-powered technologies can facilitate various forms of feedback that enhance learning and performance, particularly through automated feedback on written tasks aimed at improving writing skills.

Peer assessment involves students of similar status actively assisting and supporting each other to gain knowledge. This method has recently been recognized as an effective strategy for improving students' writing standards. However, successful outcomes from peer learning are not guaranteed without appropriate support. Traditional face-to-face peer feedback can be challenging to implement, but AI offers new opportunities for peer learning and assessment. In AI-enhanced settings, students can submit their writing and provide reciprocal, anonymous critiques of their peers' work without being constrained by time or location. These settings allow for the inclusion of various instructional tools, such as scripting and guiding, which help students deliver more reliable, accurate, and relevant feedback to their peers.

While qualitative data gives these tendencies more depth and meaning in the present study, quantitative data shows broad statistical trends. Teachers' qualitative data lends credibility to the findings by providing an explanation for their direct observations of students' performance in the two groups, which helps to explain the unexpected outcomes. Therefore, in the current study, the quantitative and qualitative data complement one another.

Leveraging AI in peer assessment can make the feedback process more efficient and effective. AI tools can provide immediate and detailed feedback, which is crucial for timely learning and improvement. Additionally, AI can help maintain consistency and objectivity in feedback, addressing the challenges posed by varying levels of expertise and subjective biases among human peers. So, it is essential to integrate AI in peer learning that not only supports the development of students' writing skills but also enhances their ability to critically assess and improve their own and others' work, preparing them for both academic and professional success.

The researchers exert control over the unexpected or conflicting outcomes during the experiment, such as the different cultural norms or educational procedures in the

two contexts. This emphasizes how crucial it is that cultural considerations be incorporated in the analysis.

However, the study acknowledges limitations such as the small sample size, which limits the generalizability of the findings. This exploratory study, with its intentionally small sample size, serves as a foundational investigation into AI-enhanced peer assessment. The findings should be interpreted as preliminary evidence, offering a conceptual framework for more extensive future research. Researchers and educators are encouraged to view these results as a starting point for further investigation rather than definitive conclusions.

The potential biases in self-reported data from surveys in addition to the shortcomings of AI use in some contexts also represent a limitation. A rigorous analysis of potential systemic biases and ethical implications is necessary when integrating AI in educational assessment. Our research revealed a number of important factors:

a) The potential for algorithmic bias:

   AI systems could unintentionally reinforce current educational disparities. The representativeness of training data has a major impact on the quality of feedback. AI models must be continuously monitored and improved.

b) Implementation of ethics suggestions:
   Establish transparent AI evaluation procedures and create continuous bias assessment systems.
   Assure human supervision and intervention skills
   Establish precise rules for the use of AI tools in educational settings.

In relation to AI feedback, despite its strength, it has a number of drawbacks, including the inability to answer complex questions, ethical issues, and technical constraints.

Future research should include developing explicit rubrics that enable alignment between human and AI assessors and providing educators with thorough AI literacy training are the first steps in a systematic approach. To further standardize evaluation techniques, baseline assessment criteria should be established. Hybrid evaluation models are the first step in the integration phase, gradually boosting AI participation while preserving crucial human supervision to guarantee moral and precise interpretations. In order to improve AI-generated feedback, continuous improvement initiatives concentrate on reviewing it frequently, integrating feedback from educators and students, and modifying implementation plans in response to continuing assessments.

The key trajectories of future research should focus on longitudinal studies that examine the sustained impact of AI-assisted peer assessment and skill development over time. Technological evolution demands investigation into emerging AI models' capabilities and the creation of adaptive feedback mechanisms. Comparative research will explore AI assessment across diverse educational levels and cultural contexts to identify best practices. Methodological advancements must emphasize sophisticated mixed-methods designs and AI bias detection frameworks, ensuring fairness and efficacy in AI-driven

educational assessments. By integrating these elements, researchers can foster ethical, transparent, and effective AI applications in education.

Another future directions for this research include expanding the sample size to include more diverse participants in order to expand the statistical analysis to support generalizability, investigating the long-term effects of AI versus human peer feedback on writing skills, and exploring the integration of other AI tools in the feedback process.

Appendix 1

Figure 1 shows the hypothetical model that clarifies the questions of the study.

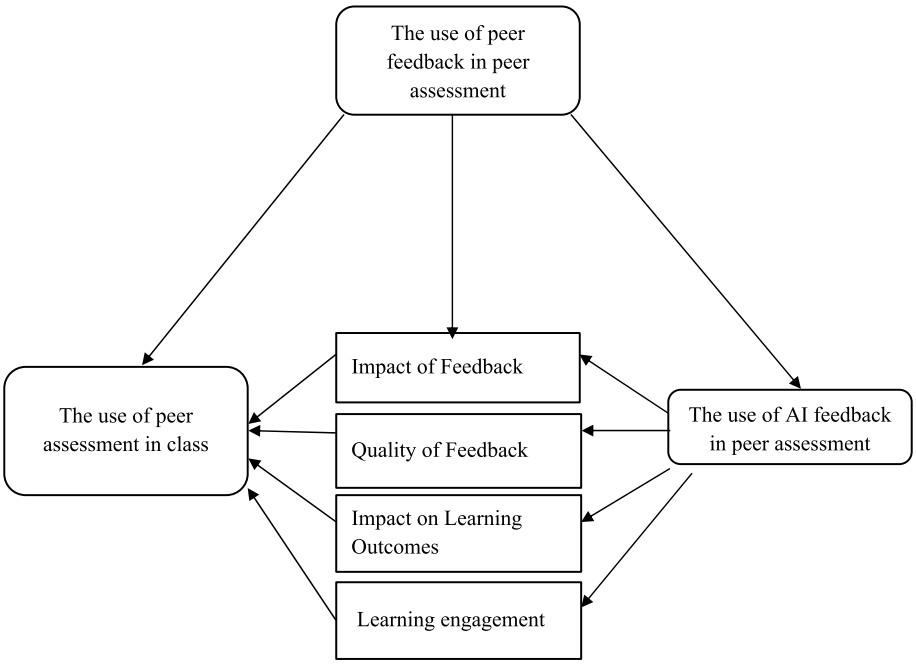

Fig. 1 The plan of the study

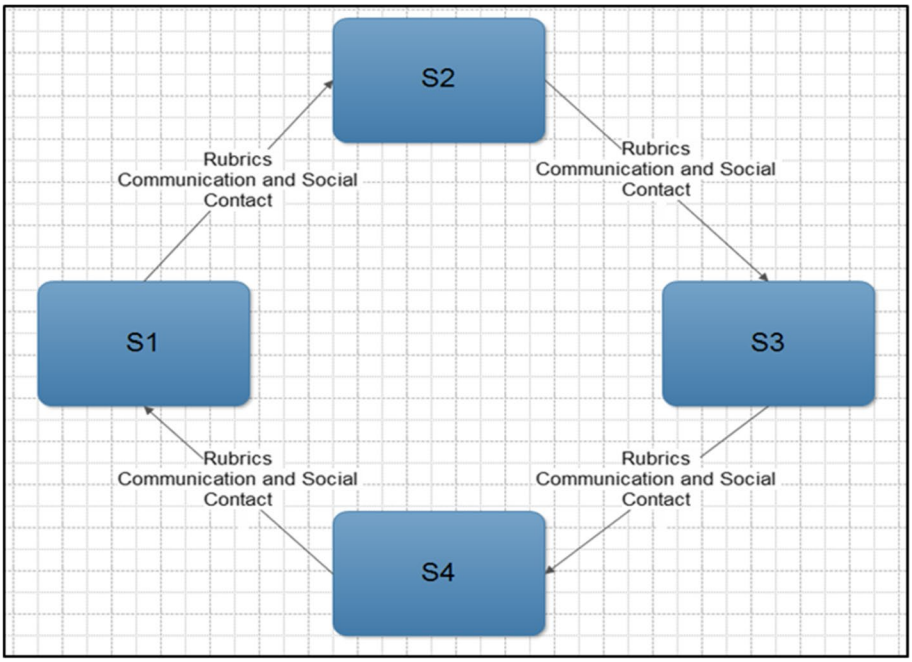

Fig. 2 Peer assessment of group A (S means student)

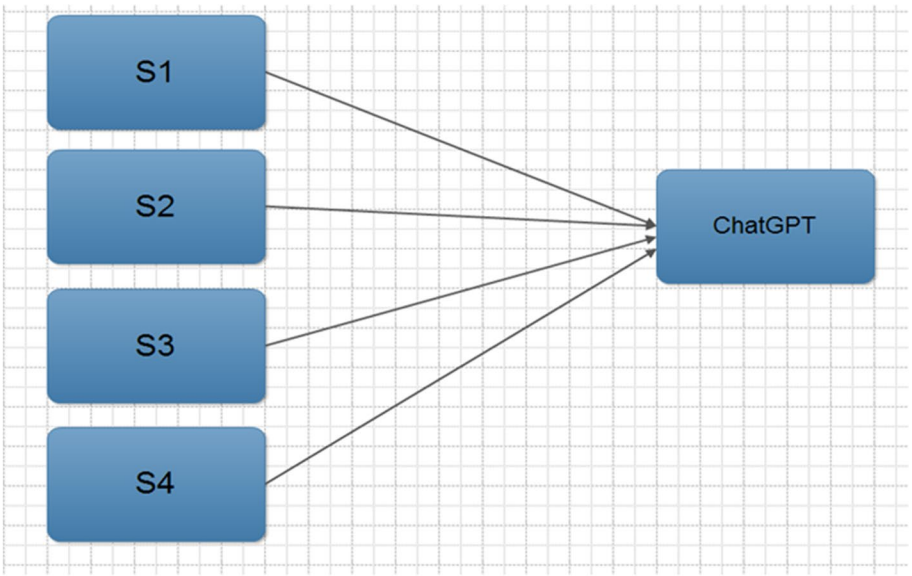

Fig. 3 Peer assessment using AI of group B (S means student)

Table 1 The detailed results gained from using the rubric in Iraq, group A

Iraq	Main idea/ focus	Organization & format	Language use & style	Originality	Creativity	Score
S1						
C1	3	2	2	1	1	7
C2	2	3	2	2	2	11
C3	3	2	2	2	3	12
C4	3	3	2	2	3	13
S2						
C1	2	2	1	2	2	9
C2	1	2	2	1	2	8
C3	1	3	3	3	1	11
C4	3	3	3	2	2	13
S3						
C1	3	3	2	2	2	12
C2	3	3	3	2	1	12
C3	2	3	1	3	2	11
C4	2	3	2	3	2	12
S4						
C1	3	2	3	2	2	12
C2	2	3	2	3	3	13
C3	3	2	2	2	3	12
C4	2	3	3	3	3	14

The detailed results gained from using the rubric in Czech Republic, group A

Czech Republic	Main idea/ focus	Organization & format	Language use & style	Originality	Creativity	Score
S1						
C1	1	3	2	1	2	9
C2	2	2	2	1	1	8
C3	2	3	1	2	2	10
C4	2	3	2	2	2	11
S2						
C1	2	2	1	2	3	10
C2	2	3	1	1	2	9
C3	1	3	3	3	3	13
C4	3	2	2	3	3	13
S3						
C1	1	2	3	2	1	9
C2	3	2	2	1	2	10
C3	3	3	3	1	2	12
C4	3	3	3	1	2	12
S4						
C1	2	2	2	1	2	9
C2	1	3	1	1	2	8
C3	3	3	2	2	3	13
C4	3	3	2	2	3	13

The detailed results gained from using ChatGPT in Iraq

Iraq	Main idea/ focus	Organization & format	Language use & style	Originality	Creativity	Score
S1						
C1	1	2	2	1	1	7
C2	1	1	2	1	1	6
C3	2	1	2	2	2	9
C4	2	2	3	2	2	11
S2						
C1	2	1	2	2	1	8
C2	1	1	1	3	2	8
C3	1	1	2	3	1	8
C4	3	2	2	3	1	11
S3						
C1	2	1	1	1	1	6
C2	1	1	1	1	2	6
C3	2	1	2	2	2	9
C4	2	1	2	2	3	10
S4						
C1	1	2	1	1	2	7
C2	2	2	1	2	2	9
C3	2	2	2	1	3	10
C4	2	3	3	2	2	12

The detailed results gained from using ChatGPT in Czech Republic

Czech Republic	Main idea/ focus	Organization & format	Language use & style	Originality	Creativity	Score
S1						
C1	1	2	2	1	1	7
C2	1	1	1	2	1	6
C3	2	2	1	1	2	8
C4	2	3	2	3	1	11
S2						
C1	2	2	1	1	1	7
C2	1	1	2	2	2	8
C3	2	1	1	2	2	8
C4	2	2	2	3	2	11
S3						
C1	1	2	1	1	1	6
C2	2	1	2	2	1	8
C3	1	3	1	2	2	9
C4	3	2	2	2	3	12
S4						
C1	2	2	1	1	2	8
C2	1	2	1	2	2	8
C3	1	2	2	2	2	9
C4	2	3	2	3	2	12

The final results of the students in the two groups

Group A Iraq Peer assessment	**Scores**	**Group B Iraq ChatGPT**	**Scores**
S1	13	S1	11
S2	13	S2	11
S3	12	S3	10
S4	14	S4	12
Group A Czech Republic Peer assessment	**Scores**	**Group B Czech Republic ChatGPT**	**Scores**
S1	11	S1	11
S2	13	S2	11
S3	12	S3	12
S4	13	S4	12

Appendix 2

The rubric for group A

Criteria	**1 Poor quality**	**2 Good quality**	**3 Excellent quality**	**Score**
Main idea/focus	The composition covers the subject poorly and contains unrelated thoughts	The composition sticks to its topic and just briefly touches on other topics	The composition exhibits a clear sense of purpose and focus, as well as insightful thoughts	
Organization & format	Insufficient or nonexistent arrangement of concepts to construct a claim	A collection of concepts arranged to construct an idea	logically and coherently constructs an idea by skillfully organizing ideas	
Language use & style	Minimal to non-existence of stylistic elements! The reader's comprehension is impeded by numerous grammatical, spelling, and punctuation errors	A few stylistic features used! Has several grammatical, spelling, and punctuation mistakes	Use of style components in an innovative and efficient way to improve meaning! Makes extremely few mistakes and consistently uses proper grammar, spelling, and punctuation	
Originality	Not a single experiment or concept improvement. Lack of commitment to the theme	A small amount of experimentation to improve ideas! Does not demonstrate originality	Creative language and usage experiments to improve ideas! Uses innovative and higher level thinking abilities to communicate difficult concepts	
Creativity	Does not demonstrate creative thinking. Does not demonstrate the ability to think creatively. Utilizes method and/or imagery that are clichéd or unoriginal	Demonstrates creative thinking. Adequate trial and error to improve ideas! Some overt allusion to the theme	Demonstrates clear exploration and ingenuity to improve imaginative notions! Obvious apparent relationship to the theme	
Total				

Appendix 3

The evaluation sheet for teachers

Iraq/Czech Republic	Group A Peer assessment		Group B ChatGpt assessment	
Criteria	Little impact The reason	High impact The reason	Little impact The reason	High impact The reason
Effectiveness of feedback				
Quality and perception of feedback				
Impact on learning outcomes				
Impact on learning engagement				
Total				

Appendix 4

The results of the two groups in evaluation sheet for teachers

Iraq/Czech Republic	Group A Peer assessment		Group B ChatGPT assessment	
Criteria	Little impact/the reason	High impact/the reason	Little impact/the reason	High impact/the reason
Effectiveness of feedback (the style)		The use of traditional peer assessment in class is very effective for the students because the process is reciprocal. The students are benefit from both when writing the composition and when revising the performance of others	Group B is a bit little than group A as the role of ChatGPT is like the role of the teacher when assessing the work	
Quality and perception of feedback	The quality of feedback in the use of peer assessment is less effective than AI for many reasons: 1) it takes much time 2) It needs much effort 3) Students' language abilities are not perfect to revise their peer's works			The quality of feedback in the use of AI as a peer assessment partner is more effective than ordinary peer assessment in class for many reasons: (1) accurate; (2) quick; (3) neutral (no passion); (4) encouraging; (5) autonomous; (6) comprehensive
Impact on learning outcomes	Learning outcomes and writing performance in group A are better because there is real process of learning; writing and revising		Learning outcomes and writing performance are less affected than group A because the role of the students is only to write a composition and receive the corrections. It is one a process. So, the process of peer assessment loss its credibility	

Iraq/Czech Republic	Group A Peer assessment		Group B ChatGPT assessment	
Criteria	**Little impact/the reason**	**High impact/the reason**	**Little impact/the reason**	**High impact/the reason**
Impact on learning engagement		Social engagement is existed more between the participants of group A. There is an exchange of discussions and objections about the final marks. Even the emotional engagement is more in group A due to the direct contact with their peers. Cognitive engagement is also found in group A because the students need to read the composition and think deeply about the existence of the criteria of assessment		Behavioral engagement is existed in the use of ChatGPT due to the increased motivation and enthusiasm of the students to learn and discover by using AI applications

Abbreviations

AI Artificial intelligence
EFL English as a Foreign Language
S Refers to the students
C Refers to composition

Acknowledgements
This research is part of the Excellence 2025 project run at the Faculty of Informatics and Management at the University of Hradec Kralove.

Authors' contributions
Both authors have prepared the manuscript and contributed equally.

Funding
No funding was received.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
The research was conducted in accordance with the Declaration of Helsinki. The research was approved by the Ethics Committee of the University of Hradec Kralove.

Consent for publication
The participants expressed their consent to take part in the research study.

Competing interests
The authors declare no competing interests.

Received: 6 October 2024 Accepted: 10 June 2025
Published: 7 July 2025

References

Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A., & Mohammed, A. A. Q. (2023). Impact of ChatGPT on learning motivation: Teachers and students' voices. *Journal of English Studies in Arabia Felix, 2*(1), 41–49. https://doi.org/10.56540/jesaf.v2i1.51

Al-Obaydi, L. H. (2021). Humanistic learning elements in a blended learning environment: A study in an EFL teaching context. *Interactive Learning Environments, 31*(5), 3098–3111. https://doi.org/10.1080/10494820.2021.1919717

Al-Obaydi, L. H., Pikhart, M., & Tawafak, R. (2023). Online assessment in language teaching environment through essays, oral discussion, and multiple-choice questions. *Computer-Assisted Language Learning Electronic Journal (CALL-EJ), 24*(2), 175–197.

Al-Obaydi, L. H., Shakki, F., & Pikhart, M. (2025). How does emergency remote teaching affect EFL students' writing skills? *Cogent Education, 12*(1), 2500001. https://doi.org/10.1080/2331186X.2025.2500001

Al-Obaydi, L. H., Shakki, F., Tawafak, R. M., Pikhart, M., & Ugla, R. L. (2023a). What I know, what I want to know, what I learned: Activating EFL college students' cognitive, behavioral, and emotional engagement through structured feedback in an online environment. *Frontiers in Psychology, 13*, 1083673. https://doi.org/10.3389/fpsyg.2022.1083673

Alqarni, T., & Alshakhi, A. (2021). The impact of negotiation as a social practice on EFL writing peer assessment sessions. *Theory and Practice in Language Studies, 11*(10), 1334–1341. https://doi.org/10.17507/tpls.1110.23

Andrade, C. (2019). Sample size and its importance in research. *Indian Journal of Psychiatry, 61*(3), 1–6. https://doi.org/10.4103/IJPSYM.IJPSYM_504_19

Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching, 53*(1), 27–31. https://doi.org/10.3200/CTCH.53.1.27-31

Andriessen, J. (2006). Arguing to learn. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 443–460). Cambridge University Press.

Babik, D., Gehringer, E., Kidd, J., Sunday, K., Tinapple, D., & Gilbert, S. (2024). A systematic review of educational online peer-review and assessment systems: Charting the landscape. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-024-10349-x

Banihashem, K., & Macfadyen, L. P. (2021). Pedagogical design: Bridging learning theory and learning analytics. *Canadian Journal of Learning and Technology, 47*(1), 1–22. https://doi.org/10.21432/cjlt27959

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education, 38*(6), 698–712. https://doi.org/10.1080/02602938.2012.691462

Braun, V., & Clarke, V. (2017). Thematic analysis. *The Journal of Positive Psychology, 12*(3), 297–298. https://doi.org/10.1080/17439760.2016.1262613

Carless, D. (2019). Feedback loops and the longer-term: Towards feedback spirals. *Assessment & Evaluation in Higher Education, 44*(5), 705–714. https://doi.org/10.1080/02602938.2018.1531108

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

Carless, D., & Winstone, N. (2020). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education, 25*(6), 1–14. https://doi.org/10.1080/13562517.2020.1782372

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access, 8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Colliot, T., Krichen, O., Girard, N., Anquetil, É., & Jamet, É. (2024). What makes tablet-based learning effective? A study of the role of real-time adaptive feedback. *British Journal of Educational Technology, 55*(3), 1–18. https://doi.org/10.1111/bjet.13439

Conijn, R., Martinez-Maldonado, R., Knight, S., Buckingham Shum, S., Van Waes, L., & van Zaanen, M. (2020). How to provide automated feedback on the writing process? A participatory approach to design writing analytics tools. *Computer Assisted Language Learning, 33*(4), 1–31. https://doi.org/10.1080/09588221.2020.1839503

Darvishi, A., Khosravi, H., Sadiq, S., & Gašević, D. (2022). Incorporating AI and learning analytics to build trustworthy peer assessment systems. *British Journal of Educational Technology, 53*(4), 844–875. https://doi.org/10.1111/bjet.13233

Darvishi, A., Khosravi, H., Sadiq, S., & Weber, B. (2021). Neurophysiological measurements in higher education: A systematic literature review. *International Journal of Artificial Intelligence in Education, 31*(2), 1–15. https://doi.org/10.1007/s40593-021-00256-0

de Boer, M., Leontjev, D., & Friederich, L. (2023). From language to function: Developing self-and peer-assessment tools. *ELT Journal, 77*(1), 94–104. https://doi.org/10.1093/elt/ccac014

Derakhshan, A., Molana, K., & Nazari, M. (2024). "I wonder who I am when my assessment ideas are not respected": A case-study inquiry into the role of institutional policies in English language teachers' assessment identity. *Asia Pacific Journal of Education, 44*(2), 1–16. https://doi.org/10.1080/02188791.2024.2328237

Derakhshan, A., & Shakki, F. (2024). The opportunities and challenges of implementing online English courses in Iranian public and private schools. *Journal of Research in Applied Linguistics, 12*(2), 1–19.

Eraut, M. (2006). Learning as dialogue, reflection, and praxis. *Journal of Workplace Learning, 18*(1), 36–50.

Gyamfi, G., Hanna, B. E., & Khosravi, H. (2021). The effects of rubrics on evaluative judgment: A randomized controlled experiment. *Assessment & Evaluation in Higher Education, 47*(1), 126–143. https://doi.org/10.1080/02602938.2021.1887081

Haßler, B., Major, L., & Hennessy, S. (2016). Tablet use in schools: A critical review of the evidence for learning outcomes. *Journal of Computer Assisted Learning, 32*(2), 139–156. https://doi.org/10.1111/jcal.12123

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

Huisman, B., Saab, N., van den Broek, P., & van Driel, J. (2019). The impact of formative peer feedback on higher education students' academic writing: A meta-analysis. *Assessment & Evaluation in Higher Education, 44*(6), 863–880. https://doi.org/10.1080/02602938.2018.1545896

Huisman, B., Saab, N., van Driel, J., & van den Broek, P. (2018). Peer feedback on academic writing: Undergraduate students' peer feedback role, peer feedback perceptions, and essay performance. *Assessment & Evaluation in Higher Education, 43*(6), 955–968.

Jansen, T., Höft, L., Bahr, L., Fleckenstein, J., Möller, J., Köller, O., & Meyer, J. (2024). Comparing generative AI and expert feedback to students' writing: Insights from student teachers. *Psychologie in Erziehung und Unterricht, 71*(2), 80–92.
Jurs, P., & Špehte, E. (2021). The role of feedback in the distance learning process. *Journal of Teacher Education for Sustainability, 23*(2), 91–105. https://doi.org/10.2478/jtes-2021-0019
Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence, 3*, Article 100074. https://doi.org/10.1016/j.caeai.2022.100074
Klimova, B., Pikhart, M., & Al-Obaydi, L. H. (2024). Exploring the potential of ChatGPT for foreign language education at the university level. *Frontiers in Psychology, 15*, 1269319. https://doi.org/10.3389/fpsyg.2024.1269319
Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(3), 1–14. https://doi.org/10.1177/00336882231162868
Latifi, S., & Noroozi, O. (2021). Supporting argumentative essay writing through an online supported peer-review script. *Innovations in Education and Teaching International, 58*(2), 195–206. https://doi.org/10.1080/14703297.2021.1961097
Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology, 41*(3), 525–536.
Lin, G. Y. (2018). Anonymous versus identified peer assessment via a Facebook-based learning application: Effects on quality of peer feedback, perceived learning, perceived fairness, and attitude toward the system. *Computers & Education, 116*, 81–92. https://doi.org/10.1016/j.compedu.2017.08.010
Lin, H., & Chen, Q. (2024). Artificial intelligence-integrated educational applications and college students' creativity and academic emotions: Students' and teachers' perceptions and attitudes. *BMC Psychology, 12*, 487. https://doi.org/10.1186/s40359-024-01979-0
Liu, N.-F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279–290.
Lu, J., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science, 40*(2), 257–275. https://doi.org/10.1007/s11251-011-9177-2
Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing, 18*(1), 30–43.
Malik, S. I., Ashfque, M. W., Tawafak, R. M., Al-Farsi, G., Usmani, N. A., & Khudayer, B. H. (2022). A chatbot to facilitate student learning in a programming 1 course: A gendered analysis. *International Journal of Virtual and Personal Learning Environments, 12*(1), 1–20.
Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S., Tondeur, J., De Laat, M., Buckingham Shum, S., Gašević, D., & Siemens, G. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI? *Computers and Education: Artificial Intelligence, 3*, Article 100056. https://doi.org/10.1016/j.caeai.2022.100056
Matcha, W., Gašević, D., & Pardo, A. (2019). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE Transactions on Learning Technologies, 13*(2), 226–245.
Moore, N. S., & MacArthur, C. A. (2016). Student use of automated essay evaluation technology during revision. *Journal of Writing Research, 8*(1), 149–175.
Nejad, A. M., & Mahfoodh, O. H. A. (2019). Assessment of oral presentations: Effectiveness of self-, peer-, and teacher assessments. *International Journal of Instruction, 12*(1), 615–632.
Noroozi, O., & Hatami, J. (2019). The effects of online peer feedback and epistemic beliefs on students' argumentation-based learning. *Innovations in Education and Teaching International, 56*(6), 548–557.
Noroozi, O., & Mulder, M. (2017). Design and evaluation of a digital module with guided peer feedback for student learning in biotechnology and molecular life sciences, attitudinal change, and satisfaction. *Biochemistry and Molecular Biology Education, 45*(1), 31–39.
Paiva, R. O., & Bittencourt, I. I. (2020). Helping teachers help their students: A human-AI hybrid approach. *Artificial Intelligence in Education, 12163*, 448–459.
Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: How students respond to peers' texts of varying quality. *Instructional Science, 43*, 591–614.
Patchan, M. M., Schunn, C. D., & Clark, R. J. (2018). Accountability in peer assessment: Examining the effects of reviewing grades on peer ratings and peer feedback. *Studies in Higher Education, 43*, 2263–2278.
Pikhart, M., & Al-Obaydi, L. H. (2025). Reporting the potential risk of using AI in higher education: Subjective perspectives of educators. *Computers in Human Behavior Reports, 7*(2), 45–62. https://doi.org/10.1016/j.chbr.2025.100693
Ranalli, J. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology, 26*(2), 5–24. https://doi.org/10.10125/73474
Reinke, W. M., Herman, K. C., & Stormont, M. (2013). Classroom-level positive behavior supports in schools implementing SW-PBIS: Identifying areas for enhancement. *Journal of Positive Behavior Interventions, 15*, 51–60.
Salman, S., & Chaya, R. (2024). The influence of AI-powered learning platforms on student engagement and performance: Emerging technologies in education. *International Journal of Research Publication and Reviews, 5*(7), 1816–1824. https://doi.org/10.55248/gengpi.5.0724.1750
Selvaraj, A. M., & Azman, H. (2020). Reframing the effectiveness of feedback in improving teaching and learning achievement. *International Journal of Evaluation and Research in Education, 9*(4), 1055–1062. https://doi.org/10.11591/ijere.v9i4.20654
Shahali Zadeh, M., Dehghani, S., Banihashem, S. K., & Rahimi, A. (2016). Designing and implementation of blending of problem-solving instructional model with constructivism principles and the study of its effect on learning and creative thinking. *Journal of Innovation and Creativity in Human Science, 5*(3), 83–117. https://journal.bpj.ir/article_522444.html?lang=en
Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL, 36*(2), 187–209. https://doi.org/10.1017/S0958344023000265
Shi, M. (2019). The effects of class size and instructional technology on student learning performance. *The International Journal of Management Education, 17*(1), 130–138. https://doi.org/10.1016/j.ijme.2019.01.004

Shoham, N., & Pitman, A. (2021). Open versus blind peer review: Is anonymity better than transparency? *Bjpsych Advances, 27*, 247–254.
Tarhini, A., Hone, K., & Liu, X. (2015). A cross-cultural examination of the impact of social, organizational, and individual factors on educational technology acceptance between British and Lebanese university students. *British Journal of Educational Technology, 46*(4), 739–755.
Tempelaar, D. (2020). Supporting the less-adaptive student: The role of learning analytics, formative assessment, and blended learning. *Assessment & Evaluation in Higher Education, 45*(4), 579–593. https://doi.org/10.1080/02602938.2019.1677855
Timans, R., Wouters, P., & Heilbron, J. (2019). Mixed methods research: What it is and what it could be. *Theory and Society, 48*, 193–216. https://doi.org/10.1007/s11186-019-09345-5
Topping, K. J. (2009). Peer assessment. *Theory into Practice, 48*, 20–27.
Topping, K. (2018). Using peer assessment to inspire reflection and learning. *Routledge*. https://doi.org/10.4324/9781351256889
Valero Haro, A., Noroozi, O., Biemans, H. J. A., & Mulder, M. (2019). The effects of an online learning environment with worked examples and peer feedback on students' argumentative essay writing and domain-specific knowledge acquisition in the field of biotechnology. *Journal of Biological Education, 53*(4), 390–398. https://doi.org/10.1080/00219266.2018.1480460
Wilson, J., Myers, M. C., & Potter, A. (2022). Investigating the promise of automated writing evaluation for supporting formative writing assessment at scale. *Assessment in Education: Principles, Policy & Practice, 29*(1), 1–17. https://doi.org/10.1080/0969594X.2022.2025762
Yang, M., Tai, M., & Lim, C. P. (2016). The role of e-portfolios in supporting productive learning. *British Journal of Educational Technology, 47*(6), 1276–1286. https://doi.org/10.1111/bjet.12412
Yu, S., & Wang, Y. (2021). Students' perceptions and preferences of generative artificial intelligence feedback for programming. *Proceedings of the AAAI Conference on Artificial Intelligence, 38*(21), 23250–23258. https://doi.org/10.1609/aaai.v38i21.30372
Yu, S., Zhang, Y., Zheng, Y., Yuan, K., & Zhang, L. (2019). Understanding student engagement with peer feedback on master's theses: A Macau study. *Assessment & Evaluation in Higher Education, 44*(1), 50–65. https://doi.org/10.1080/02602938.2018.1467879
Yüksel, D., Kaya, S., & Curle, S. (2021). What role do language learning mindsets play in English-medium instruction? A comparison of engineering and business administration in Turkey. *Journal for the Psychology of Language Learning, 3*(1), 50–62.
Zhang, H., Magooda, A., Litman, D., Correnti, R., Wang, E., Matsumura, L. C., Howe, E., & Quintana, R. (2019). eRevise: Using natural language processing to provide formative feedback on text evidence usage in student writing. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(1), 9619–9625. https://doi.org/10.1609/aaai.v33i01.33019619
AlFarsi, G., Tawafak, R. M., ElDow, A., Malik, S. I., Jabbar, J., & Al Sideiri, A. (2021). Smart classroom technology in artificial intelligence: A review paper. *International Conference on Culture Heritage, Education, Sustainable Tourism, and Innovation Technologies*, 229–235.
Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education, 21*(1), Article 23. https://doi.org/10.1186/s41239-024-00455-4
Boud, D. (2020). Challenges for reforming assessment: The next decade. *International Virtual Meeting: Teaching, Learning & Assessment in Higher Education, Universidad del Desarrollo.* https://educacionsuperior2020.udd.cl/
Bulut, O., Cutumisu, M., Singh, D., & Aquilina, A. M. (2020). Guidelines for generating effective feedback from e-assessments. *Hacettepe University Journal of Education*, 1–13. https://doi.org/10.16986/HUJE.2020063705
Creswell, J. W. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). Los Angeles: Sage Publications.
Creswell, J. W., & Plano Clark, V. L. (2017). Designing and conducting mixed methods research (3rd ed.). Los Angeles: Sage Publications.
Khaneh, M. P. A., Banihashem, S. K., & Noroozi, O. (2022). Students' motivation, satisfaction, and perceived learning using online peer feedback in the context of argumentative essay writing. In M. Shelley, H. Akcay, & O. Tayfur Ozturk (Eds.), *International Conference on Research in Education & Science* (pp. 118–121). International Society for Technology, Education, and Science (ISTES).
Khosravi, H., Gyamfi, G., Hanna, B. E., Lodge, J., & Abdi, S. (2021). Bridging the gap between theory and empirical research in evaluative judgment. *Journal of Learning Analytics, 8*(3), 117–132. https://doi.org/10.18608/jla.2021.7206
Kumar, R. (2019). *Research methodology: A step-by-step guide for beginners* (5th ed.). Sage Publications.
Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Gerardou, F. S. (2023). Challenges and opportunities of generative AI for higher education as explained by ChatGPT. *Education Sciences, 13*(9), Article 856. https://doi.org/10.3390/educsci13090856
Ortega-Ruiperez, B., & Correa-Gorospe, J. M. (2024). Peer assessment to promote self-regulated learning with technology in higher education: Systematic review for improving course design. *Frontiers in Education, 9*, Article 1376505. https://doi.org/10.3389/feduc.2024.1376505
Phuong, H.Y., Phan, Q.T. & Le, T.T. (2023) The effects of using analytical rubrics in peer and self-assessment on EFL students' writing proficiency: a Vietnamese contextual study. Lang Test Asia 13, 42. https://doi.org/10.1186/s40468-023-00256-y
Pikhart, M., Klimova, B., & Al-Obaydi, L. H. (2024). Exploring university students' preferences and satisfaction in utilizing digital tools for foreign language learning. *Frontiers in Education, 9*, Article 1412377. https://doi.org/10.3389/feduc.2024.1412377
Steiss, J., Tate, T. P., Graham, S., Cruz, J., Hebert, M., Wang, J., ... Warschauer, M. (2023). Comparing the quality of human and ChatGPT feedback on students' writing. *OSF Preprints*. https://doi.org/10.35542/osf.io/ty3em

Tawafak, R. M., Al-Obaydi, L. H., & Pikhart, M. (2023). Competency categorization and roles of online teachers from the perspective of university students. *Frontiers in Psychology, 14*, Article 1009000. https://doi.org/10.3389/fpsyg.2023.1009000

Tawafak, R. M., Al-Obaydi, L. H., Pikhart, M., & Namaziandost, E. (2024). Risk-taking, TAM model, and technology integration: Impact on EFL college students' behavioral intentions. *Applied Research on English Language, 13*(2). https://doi.org/10.22108/are.2024.140729.2241

UNESCO. (2019). Section 2: Preparing learners to thrive in the future with AI. In *Artificial intelligence in education: Challenges and opportunities for sustainable development* (pp. 17–24). United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000366994/PDF/366994eng.pdf.multi

Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology, 14*, Article 1261955. https://doi.org/10.3389/fpsyg.2023.1261955

Zhai, X. (2023). ChatGPT and AI: The game changer for education. *ResearchGate*. https://www.researchgate.net/publication/369234378_ChatGPT_and_AI_The_Game_Changer_for_Education

Publisher's Note