

Comparison of Feedback from ChatGPT and Human Professors in Higher Education: A Systematic Review

Candy Haydee Guardia-Paniura¹, Timoteo Cueva-Luza², Favio Mauricio Cruz-Carpio³, Raúl Reynaldo Ito-Díaz⁴, David Victor Apaza-Paco⁵, Nilda Rosas-Rojas⁶, Benedicta Mamani-Mamani⁴, Ángel Terrero-Pérez⁷, Renato Yassutaka Faria Yaedú^{7,8}, Mariela Peralta-Mamani^{8*}

¹School of Education, National Pedro Ruíz Gallo University, Lambayeque, Peru

²School of Psychology, Autonomous University of Ica, Ica, Peru

³School of Law, Technological University of Peru, Lima, Peru

⁴Professional School of Renewable Energy Engineering, National University of Juliaca, Juliaca, Peru

⁵School of Education, National University of San Agustín of Arequipa, Arequipa, Peru

⁶School of Education, Peruvian Union University, Puno, Peru

⁷Department of Surgery, Stomatology, Pathology and Radiology, Bauru School of Dentistry, University of São Paulo, Bauru, Brazil

⁸Hospital for Rehabilitation of Craniofacial Anomalies, University of São Paulo (HRAC-USP), Bauru, Brazil

Email: *marielaperalta@alumni.usp.br

How to cite this paper: Guardia-Paniura, C.H., Cueva-Luza, T., Cruz-Carpio, F.M., Ito-Díaz, R.R., Apaza-Paco, D.V., Rosas-Rojas, N., Mamani-Mamani, B., Terrero-Pérez, Á., Yaedú, R.Y.F. and Peralta-Mamani, M. (2025) Comparison of Feedback from ChatGPT and Human Professors in Higher Education: A Systematic Review. *Journal of Computer and Communications*, 13, 245-261.

<https://doi.org/10.4236/jcc.2025.137013>

Received: June 12, 2025

Accepted: July 25, 2025

Published: July 28, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Objectives: To compare the feedback provided by human professors and ChatGPT on university students' work and to report on students' perceptions of both types of feedback. **Materials and Methods:** A systematic review was conducted following PRISMA 2020 guidelines. Databases searched included Web of Science, SCOPUS, EBSCO, ACM Digital Library, and IEEE Xplore, with additional gray literature sources, until February 2025. Inclusion criteria were cross-sectional studies evaluating university students' work, comparing feedback from ChatGPT with human professors. Data extraction was performed using a standardized form, and risk of bias was assessed with the Joanna Briggs Institute Critical Appraisal Tool. A narrative synthesis of the results was made. PROSPERO registration number: CRD42024566691. **Results:** The review included 8 studies with 461 students. ChatGPT feedback was detailed and rapid, while human feedback was valued for its personalization and emotional support. The differences between ChatGPT and human feedback were insignificant, as both were similar. Students appreciated the detailed and immediate nature of ChatGPT feedback but noted its lack of emotional nuance and context-specific guidance. Human feedback was preferred for addressing individual learning needs and providing affective support. A combination of both types of feedback can maximize benefits. **Conclusions:** ChatGPT can assist human teachers by

providing detailed and timely feedback to university students. However, human supervision is essential to ensure feedback is nuanced and contextually appropriate. A hybrid approach can optimize the learning experience in higher education. Further research is necessary to explore AI applications in educational settings and understand their impact on learning outcomes.

Keywords

Higher Education, ChatGPT, Feedback, Systematic Review

1. Introduction

Feedback is any information given to a student after their response to inform them about their performance. Educational feedback is an effective approach to enhance student learning. However, it can be labor-intensive, which motivates the use of automated feedback tools [1] [2].

Providing feedback to higher education students is an essential skill for teachers and significantly influences the learning process. The development of writing skills in university students is crucial for their academic and professional success. Constructive feedback from teachers plays a fundamental role by offering ideas and recommendations to improve students' writing abilities. This contributes to a deeper understanding and enhances the ability to communicate effectively, having a significant impact on the professionalization of higher education [3] [4].

Providing individualized feedback to the student becomes challenging, as teachers are often overwhelmed in large classes of students. Thus, these challenges have led to looking for innovative solutions, such as automated feedback using artificial intelligence (AI) [3] [4].

ChatGPT is an intelligent AI-developed chatbot that was launched in November 2022. It has multiple applications and the ability to generate various forms of text, answer questions, and provide translations [5].

AI is a powerful data analysis tool that enhances the quality of feedback, which can boost productivity. Tools like ChatGPT can be useful for this purpose, providing individualized and timely feedback. However, they are limited in terms of quality, authenticity, and emotional intelligence. People may have a negative perception of these tools [6].

In recent years, there has been an increasing number of publications on the use of AI in education. However, there are still few studies comparing feedback from human teachers and ChatGPT, as well as student perceptions of these technologies. Student perceptions play a crucial role in motivation, engagement, and academic achievement. Therefore, the objective of this systematic review is to compare the feedback provided by human professors and ChatGPT on university students' work and to report on students' perceptions of both types of feedback.

2. Methodology

2.1. Protocol and Registration

This systematic review was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [7]. The review protocol was registered with PROSPERO, an international prospective register of systematic reviews, under the registration number CRD42024566691.

(<http://www.crd.york.ac.uk/PROSPERO/>).

2.2. Eligibility Criteria

The systematic review included studies that met the following PECO strategy: 1) Participants: university students; 2) Exposure: feedback from ChatGPT; 3) Control: feedback from human professors; and 4) Outcome: effectiveness of the feedback and student perceptions of the feedback. All cross-sectional studies that evaluated the work of university students comparing feedback from ChatGPT with feedback from human professors were included.

Studies that included postgraduate students or university professors, or that used ChatGPT for purposes other than generating feedback on university work, were excluded. Additionally, reviews and letters to the editor were excluded.

2.3. Exposure and Control

The exposure in this systematic review was feedback provided by ChatGPT, an AI chatbot. The control was feedback given by human professors. This comparison aimed to evaluate the effectiveness and perceptions of both feedback sources in the context of higher education.

2.4. Information Sources and Search

The search strategy included the following electronic bibliographic databases: Web of Science, SCOPUS, EBSCO, ACM Digital Library, and IEEE Xplore. Additionally, gray literature was searched using the Brazilian Digital Library of Theses and Dissertations (IBICT/BDTD), OpenGrey, ProQuest, and Google Scholar (first 100 records). The search terms related to “Higher education,” “ChatGPT,” and “Feedback,” and were combined using Boolean operators “OR” and “AND.” The search strategy was as follows: (“ChatGPT”) and (“higher education” OR university OR “undergraduate students” OR college OR “graduate students” OR “Post-secondary education” OR education) and (“student feedback” OR “feedback” OR “AI-generated feedback” OR “Feedback sources” OR “Peer feedback”).

No restrictions were applied to the year of publication or language initially. Additionally, a manual search was conducted to identify eligible studies. Only studies published up to February 2024 were included.

The search strategy was adapted for each database (see Supplement 1). All collected records were imported into EndNote Web (www.myendnoteweb.com), where duplicates were removed.

2.5. Study Selection

All records were imported into Rayyan software for the initial phase of study selection. In this phase, two reviewers (M.P.M. and A.T.P.) independently screened the titles and abstracts. During the second phase, the full texts of potentially eligible studies were reviewed to confirm whether they met the eligibility criteria. Any discrepancies between the reviewers were resolved by consensus.

2.6. Data Extraction and Data Items

Data extraction was performed independently by two reviewers (M.P.M. and A.T.P.) using a standardized form in Microsoft Excel. If there was any discrepancy, it was resolved by consensus. The extracted data included: first author and year of publication, geographic region, total number of participants, participant demographics (age and sex), type of work evaluated, details of the feedback provided (ChatGPT and human professors), study variables and results (**Table 1**).

Table 1. Characteristics of the included studies.

Authors and year	Geographic region	Amostra	Type of work evaluated	Exhibition and Data collection control details	Data collection time	Variables evaluated	Results
AlGhamdi, 2024	Saudi Arabia	111 first-year computing students, all men, aged 19 to 21.	Essay submitted on the "Blackboard Weekly writing assignments" blog as an integral part of the course curriculum.	Human feedback and feedback generated by ChatGPT	6 weeks. (professor feedback) 4th and 5th weeks (ChatGPT-generated feedback)	Emotional and Psychological Responses. Perceived Quality and Usefulness. Progress and Development. Content and Delivery.	109 students (1st-3rd weeks) 102 students (4th-5th weeks) Reactions vary, from the need for empathy and clarity in feedback. Feedback helps to improve and learn. Others criticize the lack of customization and consistency. It must be precise and empathetic, considering the emotional and psychological spaces of the students.
Escalante <i>et al.</i> , 2023	USA	43 undergraduate students in the Asia-Pacific region, registered in an English course, B1 proficiency level (CEFR). 13 M, 30 F, aged 19 - 36.	Writing of texts in English	Human tutors and ChatGPT-4	6 weeks. Weekly assessment	8 items, feedback, divided into 4 groups: satisfaction, clarity, usefulness, preference: 1 to 5.	18 students preferred to receive feedback from human tutors (mainly in the items of satisfaction, clarity, kindness, and preference). 20 students preferred AI. There was no significant difference. 5 students did not respond to the final survey.
Guo e Wang, 2023	China	50 undergraduate students, registered in an English course, with B2-C1 proficiency level (CEFR), 24 M, 26 F, aged 18 - 21, mean 19.54 years	300-word English text composition after class	5 human instructors (10 essays for each) and ChatGPT	1 class	Content: quality and development of arguments. Organization. Language: spelling, format	Professors paid more attention to content and language and less attention to organization. More directive and informative comments. ChatGPT provided a greater number of comments across all three evaluated parameters. More directive comments and praise.
Ivanovic, 2023	Montenegro	78 students at the Faculty of Science and Mathematics, 34 M, 44 F	English text composition	3 human professors and ChatGPT 3.5	Students had 2 months (December 2022-January 2023)	Quality of its content. Strength of arguments. The use of evidence. Relevance of the content.	Calculated ICC value of 0.8 between professors and ChatGPT indicates good consistency and reliability among evaluators.
Jukiewicz, 2024	Poland	67 students. 25 students completed 9 projects. 20 completed 8 projects, 16 completed 7 projects.	The course "Programming," in the Cognitive Science program. The number of tasks for assessment was 1579.	Human instructor and ChatGPT 3.5-turbo (each evaluated 15 tasks).	15 weeks (2x/week, 1.5 hours)	"ChatGPT Prompt Engineering for Developers" was used. The work is evaluated as correct, almost correct, or incorrect: 1, 0.5, or 0 points.	In all tasks, the average scores from the professor are slightly higher than those from ChatGPT. The standard deviation of the professor's scores is greater, indicating more variation in the ratings. ICC of the 15 responses from ChatGPT is 0.13, indicating an insignificant difference between the responses.

Continued

Lu <i>et al.</i> , 2024	China	46 education students of academic writing training program, 4 M, 42 F, 23.35 years.	A 300-word summary of a fictional article about contemporary challenges in Chinese language teaching classrooms	Human professors and ChatGPT 3.5	6 weeks (1x/week, 3 hours).	-Ability to organize content logically: five levels (0 - 8) -Express content concisely: five levels (0 - 8). Maximum total score: 40.	Moderate to good coherence between the scores of the professor and ChatGPT (ICC 0.6 to 0.75). ChatGPT provided extensive comments and general suggestions. The professors included more praise, explanations, and specific solutions.
Tossell <i>et al.</i> , 2024	USA	24 senior-year students from an engineering course at the United States Air Force Academy (USAFA) with limited experience with ChatGPT. 16 M, 8 F, 22.25 years.	Writing assignment on the current challenges of human factors and human-computer interaction.	Human instructors and ChatGPT 4.0	Approximately 2 months	- Perceived quality and difficulty: 1 - 7 - Educational value and level of comfort being responsible for the ChatGPT text: 1 - 7 - Perceived reliability: 0-7 - Confidence in the text: 0 - 7 - Confidence in the evaluation: 1 - 7 - Evaluation preference: ChatGPT, instructor, or both	- Quality: before 5.48, after 4.75 - Difficulty: before 4.8, after 5.25 - Educational value: before 5.43, after 5.57 - Responsibility: before 3.71, after 3.82 - Reliability: difference between pre and post of the ethical and benevolent subscale - Confidence in text: before 4.07, after 4.23 - Confidence in evaluation: instructor 6.29, ChatGPT 4.29, both 5.5 - Evaluation preference: 15/24 preferred instructor, 9 preferred instructor and ChatGPT
Wang <i>et al.</i> , 2024	China	42 second year students in an argumentation teaching activity	50 argumentation contents that human teachers had previously assessed. 84,000 words. Short arguments: 13 (800 - 1300 words). Medium arguments: 16 (1300 - 1800 words). Long arguments: 21 (1800 - 2300 words).	Human professors and ChatGPT 3.5	8 weeks	Evaluation dimensions: claim, evidence, rebuttal, adequacy of evidence and explanation.	-Precision rate 91.8%. Claim (100%), evidence (95.8%), and rebuttal (91.0%), the adequacy of evidence (85.3%), explanation (85%). -Recall rate 63.2%. Claim (100%), evidence (89.2%), rebuttal (75.9%), the adequacy of evidence (47.4%), explanation (29.8%)

Abbreviations: CEFR (Common European Framework of Reference for Languages), F (female), M (male), AI (Artificial Intelligence), ICC (Intraclass Correlation Coefficient).

2.7. Risk of Bias of Individual Studies

The risk of bias in individual studies was assessed using the “JBI Critical Appraisal Checklist for Analytical Cross-Sectional Studies” from the Joanna Briggs Institute [8]. This tool consists of eight items that evaluate various criteria of the studies, including the clear definition of inclusion criteria, detailed description of the subjects and the study setting, the validity and reliability of exposure measurement, the use of objective standard criteria for condition measurement, identification of confounding factors and strategies to manage them, the validity and reliability of outcome measurement, and the adequacy of the statistical analysis used. The possible responses to each item were: “Yes,” “No,” “Unclear,” or “Not applicable,” as appropriate. Two reviewers independently assessed each study, with discrepancies resolved by consensus.

2.8. Summary Measures

Effectiveness was measured using various metrics such as the intraclass coefficient-

cient, absolute values, and percentages. Student perceptions were assessed through surveys and Likert scale ratings, capturing measures of satisfaction and perceived quality of the feedback.

2.9. Synthesis of Results

A narrative synthesis of the results was structured around the comparison of feedback from ChatGPT and human professors, including student perceptions. The data synthesis focused on identifying common themes and differences in feedback effectiveness and student perceptions across the included studies.

3. Results

3.1. Study Selection

A total of five databases and gray literature were searched, yielding 288 records. After excluding duplicates, 246 records were included for the title and abstract screening phase. Of these, 210 studies were excluded as they did not meet the eligibility criteria. Out of the remaining 36 studies, 16 were deemed ineligible for this review. Therefore, 20 studies proceeded to the full-text review phase.

Thus, 12 studies were excluded for various reasons. Two studies were excluded for being reviews [1] [9], one study was excluded for using ChatGPT in the supervision of postgraduate student investigations [10], and one study focused on high school teachers' perception of ChatGPT use [11]. Another article dealt with ChatGPT feedback without comparing it to human feedback [12]. Seven studies were excluded because ChatGPT was used to assist in various tasks: a study on student perception using ChatGPT for Java programming [13], a study on using ChatGPT to create a founding team within an entrepreneurship course [14], a study on using ChatGPT to facilitate the development of educational experiences in Roblox [15], a study on the effectiveness of ChatGPT as a tool for developing English learning skills [16], a study on using ChatGPT for learning (benefits, barriers, and possible solutions) [17], a study about RECaP-GPT, which integrates human action and uses ChatGPT-4 as a feedback teaching support tool [18], and a study on using AI for a comprehensive review of existing film courses and AI-recommended courses [19]. Assim, nesta revisão sistemática foram incluídos 08 estudos [4] [20]-[26] (Figure 1).

3.2. Study Characteristics

The included studies were published between 2022 and 2024, as ChatGPT was launched in November 2022. One study was from Saudi Arabia [4], two studies were from the USA [20] [25], three studies were from China [21] [24] [26], one study was from Montenegro [22], and one study was from Poland [23].

In total, 461 higher education students were included, 202 were men and 150 women, aged 18 to 36 years. Two studies did not report the gender and age of participants [23] [26].

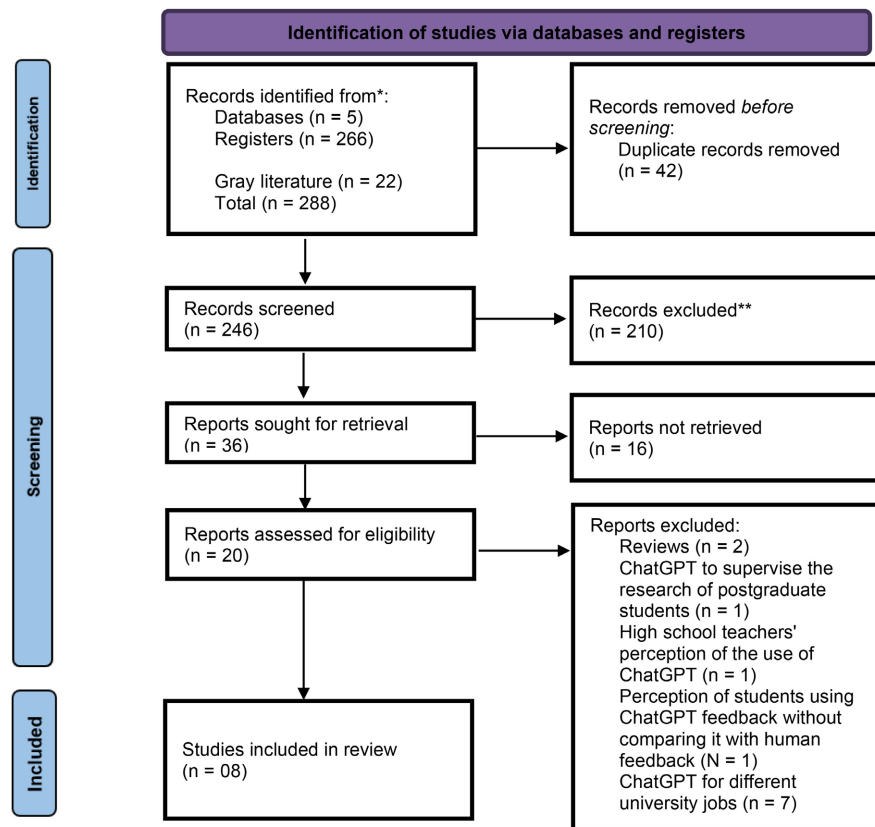


Figure 1. Flowchart of study selection for qualitative syntheses.

The evaluated work included essay writing as part of semester assignments in various fields and writing assignments in English courses. The analyzed variables included clarity, usefulness, preference, quality, organization, educational value, and confidence in the evaluation. All studies compared feedback from human instructors with feedback from ChatGPT. Two studies used ChatGPT-4 [20] [25], four studies used ChatGPT-3.5 [22]-[24] [26], and two studies did not specify the version used [4] [21].

Most of the studies evaluated student assignments over a period of 6 weeks [4] [20] [24], while others evaluated work from a single class session [21], 8 weeks [26], two months [22] [25], and 15 weeks (Table 1) [23].

3.3. Risk of Bias within Studies

The quality assessments of the individual studies were conducted using the Joanna Briggs Institute Critical Appraisal Tool for Analytical Cross-Sectional Studies. The eight included studies exhibit various levels of bias risk. Six studies present a low risk of bias due to their rigorous designs, clear inclusion criteria, and standardized evaluations [4] [20]-[22] [24] [25]. However, these studies have limitations that may affect the generalizability of the results. AlGhamdi (2024) included only male students. Tossell *et al.* (2024) used a small, homogeneous sample of USAFA cadets. Escalante *et al.* (2023) had self-selected participants and a homo-

geneous sample. Guo and Wang (2023) worked with a limited sample of five teachers and self-selected participants. Ivanović (2022) faced limitations due to the limited variability in the sample and the possible influence of human evaluators. Lu *et al.* (2024) presented a limitation in the homogeneity of the sample of Chinese students and the possible influence of the feedback sequence.

On the other hand, the studies by Jukiewicz (2024) and Wang *et al.* (2024) present a moderate risk of bias due to the lack of detailed information about the age and gender of the participants. Although both studies used blind designs and standardized evaluations, the homogeneity of their samples limits the generalizability and representativeness of the results (Figure 2).

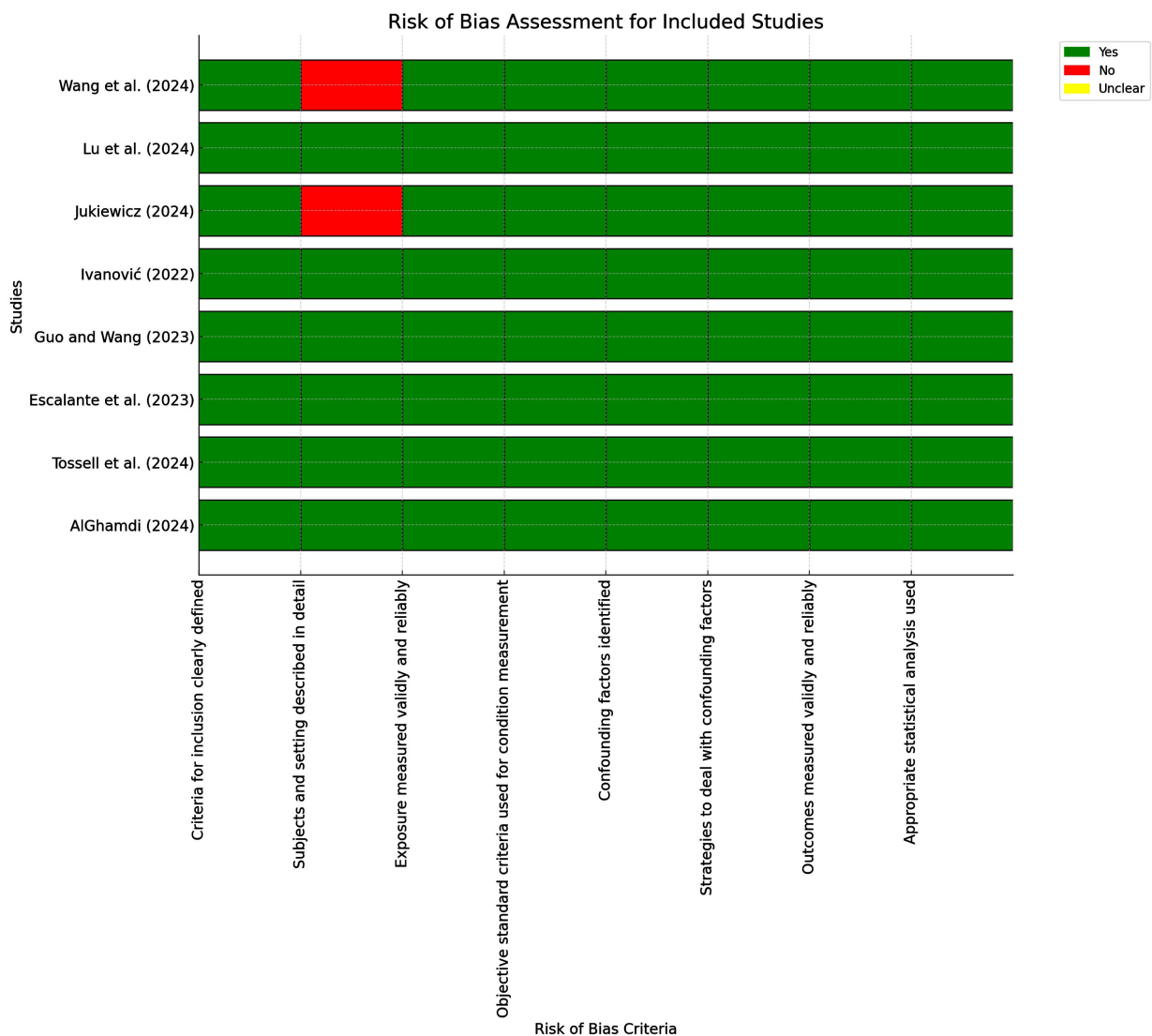


Figure 2. Risk of bias graph.

3.4. Results of Individual Studies

The study conducted by AlGhamdi (2024) found that feedback generated by

ChatGPT had emotional, psychological, and educational impacts on first-year computing students. Responses to the feedback ranged from positive emotions, such as motivation and enthusiasm, to negative ones, such as frustration and confusion. Regarding quality and usefulness, some students appreciated the detailed improvements provided by ChatGPT, while others criticized its lack of consistency and personalization. In terms of development and progress, many students acknowledged improvements in their writing skills due to regular and detailed feedback, although some noted the lack of personalization compared to human feedback. In summary, the study highlights the potential of ChatGPT to provide useful and timely feedback but emphasizes the need to complement it with human comments to more effectively address the emotional and educational needs of students.

The study by Escalante *et al.* (2023) indicates that there were no significant differences between the feedback generated by ChatGPT-4 and human instructors. Approximately the same number of students preferred AI-generated feedback and human feedback. Some characteristics of AI feedback include clarity and specificity, while human feedback is valued for its affective benefits and direct interaction. The results suggest that AI-generated feedback can be incorporated into student essay evaluations without negatively affecting learning outcomes, and they recommend a mixed approach that combines the strengths of both types of feedback.

The results of the study by Guo and Wang (2023) showed that ChatGPT generated longer, more detailed, and specific feedback compared to human instructors, who focused on issues related to content and language. Additionally, ChatGPT provided more balanced comments. Instructors expressed both positive and negative perceptions, noting that ChatGPT can complement their own feedback. However, human supervision and adjustment are necessary to maximize its effectiveness in developing writing skills.

The study by Ivanovic (2023) compared feedback from human instructors and ChatGPT, finding good consistency and reliability (Intraclass Correlation Coefficient ICC of 0.8). ChatGPT can evaluate work in less than 30 seconds and provide detailed analysis similar to human instructors, objectively based on training data. It may be more lenient with minor errors, often giving slightly higher grades. However, it is not capable of capturing emotional and cultural nuances and has difficulty detecting inconsistencies in lengthy texts. On the other hand, human instructors take about 10 to 30 minutes to perform evaluations, providing detailed and individualized analysis. However, they can be more critical and may be influenced by personal biases and fatigue, potentially resulting in lower grades. ChatGPT can serve as an evaluation assistant, offering immediate and meaningful feedback, thereby reducing the workload.

The study by Jukiewicz (2024) found a strong positive correlation between grades given by ChatGPT and human instructors, with an insignificant difference between the two. ChatGPT-generated grades were slightly lower than those given by human instructors, as ChatGPT appeared to be stricter regarding program-

ming assignment standards and more adept at detecting code errors. Human instructors tended to give higher grades to work that wasn't perfect, provided the code functioned and met the task requirements. This study evaluated assignments in a Programming course within the cognitive science program, using Python programming tasks.

Lu *et al.* (2024) evaluated the differences in feedback provided by ChatGPT and human instructors for academic writing tasks in Chinese. They found moderate to good consistency between the scores given by human instructors and ChatGPT (ICC 0.6 and 0.75). ChatGPT provided more general and extensive evaluations, while human instructors offered specific explanations and solutions. Human instructor feedback was more frequently implemented by students (80.2%) compared to ChatGPT feedback (59.9%). The integration of ChatGPT in evaluations promoted a deeper understanding and independent thinking in student revisions, significantly improving their academic writing.

The study by Tossell *et al.* (2024) indicates that ChatGPT did not simplify students' writing tasks but changed how they perceive and approach assignments given by instructors, thereby improving their learning. Initially, students viewed ChatGPT as a fraudulent tool requiring human supervision, technical competence, and calibrated trust. After using it, students recognized it as a valuable learning tool, perceiving it as more ethical and benevolent. Despite this, they showed low comfort in taking responsibility for tasks completed with ChatGPT's assistance due to ethical concerns and a lack of confidence in the accuracy of its results. Students preferred to be evaluated by both ChatGPT and the instructor, rather than by ChatGPT alone.

The study by Wang *et al.* (2024) evaluated ChatGPT's ability to provide feedback on university students' arguments and found that ChatGPT demonstrated high accuracy (91.8%) in evaluating quantitative points such as claims, evidence, and refutations, although its recall rate was 63.2%. ChatGPT's accuracy decreased with longer arguments and was influenced by the use of discourse markers. It provides more extensive, rapid, and text-based feedback, relying on data, but struggles to deliver affective feedback that is effective for students. In contrast, feedback from human instructors is more focused and based on experience.

4. Discussion

Chatbots have garnered considerable interest among academics and educators, particularly since their release in 2022. ChatGPT can contribute in various ways to the educational context, serving as a language learning tool, significantly enhancing linguistic skills, and boosting students' motivation and autonomy. This enables students to take control of their own learning process. Additionally, it can be used to generate feedback on university assignments [5]. O ChatGPT também pode abordar desafios cognitivos, não cognitivos e metacognitivos enfrentados por alunos em Ambientes de Aprendizagem Pessoal (PLEs), fornecendo informações relevantes e confiáveis, recursos de aprendizagem personalizados e

promovendo atividades de pensamento crítico [27].

Feedback is essential for improving students' written communication skills, guiding their development and refinement [4]. Students positively receive feedback from teachers, which contributes to the self-regulation of their learning. AI-provided feedback, such as that from ChatGPT, can generate complex emotional responses and affect the perception of quality and the development of the writing process. In particular, ChatGPT can significantly enhance university students' writing skills by promoting self-reflection and evaluation of their strengths and weaknesses [4] [25]. These activities not only develop written communication skills but also prepare students for the professional market.

Recent research has evaluated the efficacy of ChatGPT in education, highlighting its ability to provide comprehensive, coherent, immediate, and personalized feedback. Some studies show that ChatGPT can outperform human instructors by offering much more detailed feedback, which fluently summarizes students' performance, thereby significantly contributing to the improvement of their learning skills [2].

A study involving computer science university students asked them to develop a 2500-word report integrating personal reflections and theoretical learning. They shared their perceptions of the feedback received from both ChatGPT and human professors. Although the overall perception was positive, some students found ChatGPT's comments unclear [4]. Using ChatGPT for feedback on assignments could be a viable alternative, provided it is balanced with a personalized human touch.

In the study by Tossell *et al.* (2024), students used ChatGPT to generate and edit texts with automated feedback and comments from professors. The majority (15 out of 24) preferred a combination of instructor and ChatGPT, while 9 chose only the instructor, demonstrating the effectiveness of combining AI and human input to improve writing. ChatGPT was seen as a valuable collaborative resource under human supervision, not as a tool for cheating. Another study by Lu *et al.* (2024) compared feedback from professors and ChatGPT without students knowing the source. The results showed differences in students' understanding and preferences based on their writing level, suggesting that both types of feedback can complement each other. Together, the studies indicate that combining AI and human feedback can enhance the learning experience and improve students' writing.

A systematic review of automated grading and feedback tools in education analyzed 121 articles from 2017 to 2021, categorizing them by the skills assessed, approach, language paradigm, degree of automation, and evaluation techniques. Most studies focused on grading assignments in object-oriented languages using dynamic unit tests or static analysis techniques. These tools offer automated evaluation, instant feedback, and allow for multiple submissions, increasing student satisfaction [28].

Another study compared the feedback from five teachers and ChatGPT, finding

that ChatGPT offered flexible and personalized comments, but sometimes these were too lengthy and irrelevant, causing anxiety and stress among students with low English proficiency. This hindered their understanding and acceptance. The feedback did not directly point out errors, making it difficult to locate problems. Despite these perceptions, the teachers acknowledged that ChatGPT could complement their comments and reduce their workload [21].

A study using ChatGPT-4 to provide feedback on university students' English writing found no significant difference between ChatGPT and human professors in terms of linguistic progress. Some students preferred human feedback for the personal interaction and the ability to ask questions, while others valued the precision and continuous availability of AI feedback [20]. ChatGPT, though more lenient in grading essays and devoid of emotions or biases, relies on algorithms that analyze grammar, syntax, coherence, and relevance [22]. This suggests that while AI is useful in evaluation, it should be combined with human supervision to ensure balanced and fair assessment.

Feedback acts as a compass, guiding learning and fostering introspection and creativity. Although it can be overwhelming, it must be precise and empathetic, considering the emotional and psychological aspects of students. Teachers should provide enriching and empowering feedback that fosters confidence [4]. While students appreciate its benefits, some criticize the lack of personalization and consistency, highlighting the need for a balanced approach between human and AI feedback to optimize educational outcomes and address the emotional needs of students.

ChatGPT offers instant, objective, and extensive feedback, whereas teachers provide more limited but focused and human feedback, using affective and motivational phrases that highlight students' effort and improvement, fostering confidence and self-efficacy [26]. These limitations may stem from time and experience constraints. A study revealed that feedback can motivate, increase confidence, and promote self-reflection and growth, emphasizing the need for empathy and clarity in communication [4]. The interplay between emotion and learning contributes to knowledge, reflection, and opportunities for academic and personal improvement.

Regarding time optimization, ChatGPT takes only 9.5 to 30 seconds to generate feedback on university assignments [22] [23], saving teachers significant time, especially when analyzing large amounts of data. In contrast, a teacher may take between 10 and 30 minutes to evaluate an essay, depending on its length. This AI-assisted approach represents a paradigm shift in education, improving the efficiency and quality of assessment, benefiting both teachers and students [22]. Although ChatGPT is a useful tool, it should not be used alone. Evaluations should be complemented with other tools and manual analysis.

Two studies compared the grades given to student assignments by ChatGPT and human teachers. Ivanović (2022) found that ChatGPT tends to be more lenient, awarding higher grades to student essays due to its focus on grammar, syntax,

coherence, and relevance. Conversely, Jukiewicz (2024) discovered that ChatGPT gave lower grades on Python programming tasks, being stricter in applying programming standards and detecting errors in the code. This suggests that ChatGPT is more forgiving of minor errors in written essays but more stringent in evaluating code quality in programming tasks. On the other hand, teachers tend to give higher grades to imperfect work due to errors such as lack of attention, fatigue, or grade inflation.

Upon receiving feedback, students should review it to improve, incorporate new ideas, and monitor their progress. Future research should explore adaptive support measures based on natural language processing, considering challenges such as data scarcity and bias, and promoting interdisciplinary collaboration to develop effective automated educational support [1].

AI, specifically ChatGPT, can optimize research supervision without replacing human supervisors. ChatGPT facilitates self-directed learning and provides formative feedback, aiding in the formulation of questions and the search for evidence [9]. A study showed that its use improves research quality, fosters critical thinking, and accelerates progress, allowing supervisors to focus on strategic direction while students gain autonomy. ChatGPT promotes a shift in supervisory roles, where supervisors provide strategic guidance through periodic meetings, transforming students into autonomous researchers supported by educational AI [10]. However, AI can make mistakes and lacks a specific ethical framework, so universities should develop AI literacy protocols for responsible use, addressing concerns about dependency, intellectual development, and ethics.

The use of AI algorithms in the acquisition and processing of video stimuli significantly enhances the efficiency of stimulated recall in education. AI facilitates the automatic selection of relevant interactions and the extraction of important information, reducing cognitive load and analysis time. It can also conduct interviews, optimizing resources and increasing participation. However, it raises ethical and privacy challenges [29].

There are concerns about the long-term implications of using AI in education, including ethical issues, academic integrity, accuracy, and impacts on personal development (Chan & Hu, 2023; Firat, 2023). Despite these concerns, integrating AI tools like ChatGPT offers promising possibilities for enhancing education and complementing traditional teaching methods [4]. ChatGPT has valuable potential in education, improving outcomes and aiding student motivation, although it cannot replace human teachers [11]. Using ChatGPT to generate feedback for large groups appears to be an excellent tool. However, human supervision is essential to address emotional aspects, as feedback is not always clear and can lead to a negative experience for the student.

To integrate AI tools into teaching, clear pedagogical objectives must be defined. It is essential to introduce them gradually and ensure the necessary technical support. ChatGPT has limitations, such as potential biases and difficulties in capturing complex language nuances [22]. Universities should focus on integrating AI tools into teacher training programs, as educators need adequate training

to enhance their skills in using ChatGPT as a teaching and learning tool.

This systematic review has several limitations that may influence the interpretation of its findings. First, the included studies exhibited sample homogeneity in terms of educational level, geographic region, and discipline, limiting the generalizability of the conclusions across diverse academic contexts. Second, the rapid development and frequent updates of AI models, such as ChatGPT, may render some results outdated or less applicable to newer versions, impacting the temporal validity of the findings. Third, emotional and contextual nuances, key elements in effective pedagogical feedback, are difficult to assess and quantify in AI-generated responses, which may influence students' perceptions and affective engagement. These limitations suggest that while the findings offer valuable insights into the comparative performance of human and AI feedback, caution is warranted when extrapolating them to broader educational scenarios or future AI tools.

It is recommended to investigate the long-term dynamics between students and AI technology. Future research should focus on longitudinal studies to better understand the impacts of AI-assisted feedback on student learning outcomes over time. There is a need to explore diverse educational settings and disciplines to evaluate the broader applicability of these findings. It is crucial to examine the ethical implications and potential biases inherent in AI tools, ensuring they complement rather than replace the human elements essential to education. While these tools can enhance efficiency and access to knowledge, it is important to preserve students' intellectual autonomy and address potential risks of dependency. Recognizing and managing the ethical benefits and challenges of using AI in education is fundamental. In summary, the findings indicate that technologies like ChatGPT do not eliminate the need for student and teacher participation but rather complement it, requiring a judicious combination of human skills and AI capabilities.

5. Conclusion

ChatGPT provides effective and timely feedback, enhancing learning through detailed and rapid responses. However, its lack of emotional nuance and specific guidance suggests it cannot completely replace human feedback in higher education. Integrating feedback from ChatGPT with that from human professors can optimize the learning experience. A hybrid approach that combines both forms of feedback may be the most effective strategy for improving educational outcomes in higher education. Students value the speed and detail of ChatGPT's feedback but prefer the personalization and empathy of feedback from human professors. Further studies are needed to explore the various applications of AI in teaching university students.

Acknowledgements

This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES), "Bolsista CAPES/BRASIL" Finance Code nº

88887.967267/2024-00.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Bauer, E., Greisel, M., Kuznetsov, I., Berndt, M., Kollar, I., Dresel, M., *et al.* (2023) Using Natural Language Processing to Support Peer-Feedback in the Age of Artificial Intelligence: A Cross-Disciplinary Framework and a Research Agenda. *British Journal of Educational Technology*, **54**, 1222-1245. <https://doi.org/10.1111/bjet.13336>
- [2] Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y., Gašević, D., *et al.* (2023) Can Large Language Models Provide Feedback to Students? A Case Study on ChatGPT. 2023 *IEEE International Conference on Advanced Learning Technologies (ICALT)*, Orem, 10-13 July 2023, 323-325. <https://doi.org/10.1109/icalt58122.2023.00100>
- [3] Al-Bashir, M., Kabir, M. and Rahman, I. (2016) The Value and Effectiveness of Feedback in Improving Students' Learning and Professionalizing Teaching in Higher Education. *Journal of Education and Practice*, **7**, 38-41.
- [4] Al-Ghamdi, R. (2024) Exploring the Impact of ChatGPT-Generated Feedback on Technical Writing Skills of Computing Students: A Blinded Study. *Education and Information Technologies*, **29**, 18901-18926. <https://doi.org/10.1007/s10639-024-12594-2>
- [5] Xiao, Y. and Zhi, Y. (2023) An Exploratory Study of EFL Learners' Use of ChatGPT for Language Learning Tasks: Experience and Perceptions. *Languages*, **8**, Article 212. <https://doi.org/10.3390/languages8030212>
- [6] Tong, S., Jia, N., Luo, X. and Fang, Z. (2021) The Janus Face of Artificial Intelligence Feedback: Deployment versus Disclosure Effects on Employee Performance. *Strategic Management Journal*, **42**, 1600-1631. <https://doi.org/10.1002/smj.3322>
- [7] Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., *et al.* (2021) The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ*, **372**, n71. <https://doi.org/10.1136/bmj.n71>
- [8] Moola, S., Munn, Z., Tufanaru, C., Aromataris, E., Sears, K., Sfetic, R., *et al.* (2019) Chapter 7: Systematic Reviews of Etiology and Risk. In: Aromataris, E. and Munn, Z., Eds., *JBI Reviewer's Manual*, JBI, 219-271. <https://doi.org/10.46658/jbirm-17-06>
- [9] Cowling, M., Crawford, J., Allen, K. and Wehmeyer, M. (2023) Using Leadership to Leverage ChatGPT and Artificial Intelligence for Undergraduate and Postgraduate Research Supervision. *Australasian Journal of Educational Technology*, **39**, 89-103. <https://doi.org/10.14742/ajet.8598>
- [10] Dai, Y., Lai, S., Lim, C.P. and Liu, A. (2023) ChatGPT and Its Impact on Research Supervision: Insights from Australian Postgraduate Research Students. *Australasian Journal of Educational Technology*, **39**, 74-88. <https://doi.org/10.14742/ajet.8843>
- [11] El-Sayary, A. (2023) An Investigation of Teachers' Perceptions of Using ChatGPT as a Supporting Tool for Teaching and Learning in the Digital Era. *Journal of Computer Assisted Learning*, **40**, 931-945. <https://doi.org/10.1111/jcal.12926>
- [12] Yan, D. (2024) Feedback Seeking Abilities of L2 Writers Using ChatGPT: A Mixed Method Multiple Case Study. *Kybernetes*, **54**, 3757-3781. <https://doi.org/10.1108/k-09-2023-1933>
- [13] Haindl, P. and Weinberger, G. (2024) Students' Experiences of Using ChatGPT in an

- Undergraduate Programming Course. *IEEE Access*, **12**, 43519-43529. <https://doi.org/10.1109/access.2024.3380909>
- [14] Hammada, B. (2024) ChatGPT for Founding Teams: An Entrepreneurial Pedagogical Innovation. *International Journal of Technology in Education*, **7**, 154-173. <https://doi.org/10.46328/ijte.530>
- [15] Ho, W. and Lee, D. (2023) Enhancing Engineering Education in the Roblox Metaverse: Utilizing ChatGPT for Game Development for Electrical Machine Course. *International Journal on Advanced Science, Engineering and Information Technology*, **13**, 1052-1058. <https://doi.org/10.18517/ijaseit.13.3.18458>
- [16] Muniandy, J. and Selvanathan, M. (2024) ChatGPT, a Partnering Tool to Improve ESL Learners' Speaking Skills: Case Study in a Public University, Malaysia. *Teaching Public Administration*, **43**, 4-20. <https://doi.org/10.1177/01447394241230152>
- [17] Ngo, T.T.A. (2023) The Perception by University Students of the Use of ChatGPT in Education. *International Journal of Emerging Technologies in Learning (ijET)*, **18**, 4-19. <https://doi.org/10.3991/ijet.v18i17.39019>
- [18] Ossa, C. and Willatt, C. (2023) Providing Academic Writing Feedback Assisted by Generative Artificial Intelligence in Initial Teacher Education Contexts. *European Journal of Education and Psychology*, **16**, 1-16.
- [19] Yang, W., Lee, H., Wu, R., Zhang, R. and Pan, Y. (2023) Using an Artificial-Intelligence-Generated Program for Positive Efficiency in Filmmaking Education: Insights from Experts and Students. *Electronics*, **12**, Article 4813. <https://doi.org/10.3390/electronics12234813>
- [20] Escalante, J., Pack, A. and Barrett, A. (2023) AI-Generated Feedback on Writing: Insights into Efficacy and ENL Student Preference. *International Journal of Educational Technology in Higher Education*, **20**, Article No. 57. <https://doi.org/10.1186/s41239-023-00425-2>
- [21] Guo, K. and Wang, D. (2023) To Resist It or to Embrace It? Examining ChatGPT's Potential to Support Teacher Feedback in EFL Writing. *Education and Information Technologies*, **29**, 8435-8463. <https://doi.org/10.1007/s10639-023-12146-0>
- [22] Ivanovic, I. (2023) Can AI-Assisted Essay Assessment Support Teachers? A Cross-sectional Mixed-Methods Research Conducted at the University of Montenegro. *Annales-Analiza Istrske in Mediteranske Studije-Series Historia et Sociologia*, **33**, 571-590.
- [23] Jukiewicz, M. (2024) The Future of Grading Programming Assignments in Education: The Role of ChatGPT in Automating the Assessment and Feedback Process. *Thinking Skills and Creativity*, **52**, Article ID: 101522. <https://doi.org/10.1016/j.tsc.2024.101522>
- [24] Lu, Q., Yao, Y., Xiao, L., Yuan, M., Wang, J. and Zhu, X. (2024) Can ChatGPT Effectively Complement Teacher Assessment of Undergraduate Students' Academic Writing? *Assessment & Evaluation in Higher Education*, **49**, 616-633. <https://doi.org/10.1080/02602938.2024.2301722>
- [25] Tossell, C.C., Tenhundfeld, N.L., Momen, A., Cooley, K. and de Visser, E.J. (2024) Student Perceptions of ChatGPT Use in a College Essay Assignment: Implications for Learning, Grading, and Trust in Artificial Intelligence. *IEEE Transactions on Learning Technologies*, **17**, 1069-1081. <https://doi.org/10.1109/tlt.2024.3355015>
- [26] Wang, L., Chen, X., Wang, C., Xu, L., Shadiev, R. and Li, Y. (2024) ChatGPT's Capabilities in Providing Feedback on Undergraduate Students' Argumentation: A Case Study. *Thinking Skills and Creativity*, **51**, Article ID: 101440. <https://doi.org/10.1016/j.tsc.2023.101440>

- [27] Xu, X., Wang, X., Zhang, Y. and Zheng, R. (2024) Applying ChatGPT to Tackle the Side Effects of Personal Learning Environments from Learner and Learning Perspective: An Interview of Experts in Higher Education. *PLOS ONE*, **19**, e0295646. <https://doi.org/10.1371/journal.pone.0295646>
- [28] Messer, M., Brown, N.C.C., Kölling, M. and Shi, M. (2024) Automated Grading and Feedback Tools for Programming Education: A Systematic Review. *ACM Transactions on Computing Education*, **24**, 1-43. <https://doi.org/10.1145/3636515>
- [29] Zhai, X., Chu, X., Wang, M., Tsai, C.C., Liang, J.C. and Spector, J.M. (2024) A Systematic Review of Stimulated Recall (SR) in Educational Research from 2012 to 2022. *Humanities and Social Sciences Communications*, **11**, Article No. 489.