Medición y Evaluación del Aprendizaje Inclusivo. 2025; 2(2):14
doi: 10.63688/aprendizaje.v2.i2.14

ORIGINAL

Effect of formative feedback with artificial intelligence on learning and accessibility of university assessment

Efecto de la retroalimentación formativa con inteligencia artificial sobre el aprendizaje y la accesibilidad de la evaluación universitaria

Darwin Joel Lozada López[1]

[1] Centro de Investigación, Desarrollo e Innovación JYMNIJA Oruro - Bolivia.

Submitted: 23-08-2025 | Revised: 16-09-2025 | Accepted: 03-11-2025 | Published: 05-11-2025

Corresponding Author: Darwin Joel Lozada López

ABSTRACT

Introduction: Generative artificial intelligence offers new possibilities to strengthen formative feedback and improve the accessibility of evaluation processes in higher education. Objective: To evaluate the effect of AI-mediated formative feedback on academic performance and the perception of clarity, usefulness, accessibility, and equity of assessment in university students. Methodology: A quantitative, explanatory, and quasi-experimental study was developed with a pre-test-posttest design and non-equivalent groups. A total of 128 students participated, divided into an experimental group and a control group of 64 participants each. The intervention was developed over eight weeks and the effect on performance was analyzed by controlling the initial score. Results: The experimental group increased its mean score from 59.8 to 79.6 points, while the control group went from 60.1 to 69.1. Likewise, the experimental group obtained higher scores in clarity (4.47), utility (4.51), accessibility (4.43), and equity (4.32). The advantage was greater among students with low initial performance. Conclusion: Feedback mediated by artificial intelligence showed potential to improve learning and favor a more accessible, useful, and inclusive evaluation experience when integrated with teacher supervision.

Keywords: artificial intelligence; formative assessment; feedback; inclusive education; academic performance.

RESUMEN

Introducción: La inteligencia artificial generativa ofrece nuevas posibilidades para fortalecer la retroalimentación formativa y mejorar la accesibilidad de los procesos de evaluación en educación superior. Objetivo: Evaluar el efecto de la retroalimentación formativa mediada por inteligencia artificial sobre el rendimiento académico y la percepción de claridad, utilidad, accesibilidad y equidad de la evaluación en estudiantes universitarios. Metodología: Se desarrolló un estudio cuantitativo, explicativo y cuasiexperimental con diseño pretest–postest y grupos no equivalentes. Participaron 128 estudiantes, distribuidos en un grupo experimental y un grupo control de 64 participantes cada uno. La intervención se desarrolló durante ocho semanas y el efecto sobre el rendimiento se analizó controlando la puntuación inicial. Resultados: El grupo experimental incrementó su puntuación media de 59,8 a 79,6 puntos, mientras que el grupo control pasó de 60,1 a 69,1. Asimismo, el grupo experimental obtuvo puntuaciones superiores en claridad (4,47), utilidad (4,51), accesibilidad (4,43) y equidad (4,32). La ventaja fue mayor entre estudiantes con rendimiento inicial bajo. Conclusión: La retroalimentación mediada por inteligencia artificial mostró potencial para mejorar el aprendizaje y favorecer una experiencia evaluativa más accesible, útil e inclusiva cuando se integra con supervisión docente.

Palabras clave: inteligencia artificial; evaluación formativa; retroalimentación; educación inclusiva; rendimiento académico.

INTRODUCTION

The use of generative artificial intelligence in higher education has been gradually changing the ways in which content is produced, academic support is provided, and learning assessment processes are developed. Its deployment has shifted the discussion from the instrumental use of digital tools toward issues related to the quality of assessment, the validity of learning evidence, and the capacity of these technologies to provide responses tailored to students' needs. The most recent reviews indicate that among the applications that have grown the most is the automated generation of comments, explanations, and guidance after assessment tasks, especially in formative assessment processes [1,2]. However, the speed of technology adoption in various scenarios has been greater than the generation of empirical evidence to establish under what conditions these tools actually favor learning and when they can bring new obstacles to assessment processes [3].

Formative assessment does not pursue the purpose of giving a grade, but rather aims to produce information that helps students recognize errors, understand quality criteria, and make adjustments during the learning process. In this context, generative artificial intelligence offers potentially positive features, including the ability to provide immediate responses, develop personalized explanations, and remain available regardless of the number of students served. The evidence accumulated through experimental studies and meta-analyses indicates that the educational use of generative tools may be associated with improvements in certain learning outcomes; however, the magnitude of these effects varies according to pedagogical purpose, intervention design, discipline, exposure time, and the way technology is integrated into academic activities [4,5]. Therefore, the positive results attributed to artificial intelligence cannot be automatically transferred to any assessment modality.

In this process, feedback is among the central mechanisms through which assessment can generate learning. Its usefulness depends not only on the amount of information provided, but also on the student's ability to understand, interpret, and convert such information into subsequent actions. Indeed, the concept of information literacy has highlighted students' ability to recognize the value of the information they receive, form judgments about their own performance, and use that information to improve in future tasks [6]. More recent research has expanded this approach through instruments aimed at measuring behaviors that evidence effective student engagement with feedback [7]. From this point of view, the use of artificial intelligence becomes important when it allows a dynamic interaction with feedback, rather than when it operates only as an automatic correction system.

Generative artificial intelligence can modify this relationship because it allows requesting clarifications, reformulating explanations, and generating new examples in an iterative manner. Therefore, it has been proposed that these tools can act as facilitators of student participation in feedback processes, provided that their use is integrated within a pedagogical structure that preserves the student's capacity for judgment and academic supervision [8]. This characteristic differentiates artificial intelligence-mediated feedback from traditional models in which the comment is usually unidirectional and delivered after a given activity has been completed. However, having greater opportunities for interaction does not necessarily imply that the generated information has the same accuracy, relevance, or usefulness as that provided by a teacher.

Comparative studies developed recently show disparate results. Differences have been found between AI-generated feedback and that provided by instructors regarding specificity, length, timeliness, perceived usefulness, and student willingness to use the received comments [9]. Other works have pointed out that automation can increase the speed and consistency of feedback, although its effectiveness continues to depend on the structure of the instructions, the nature of the task, and the supervision of the produced response [10]. Similarly, comparisons between evaluations carried out by teachers, peers, and generative systems show that no feedback source presents uniform superiority across all evaluated dimensions [11]. These differences suggest that the relevant question is not only whether artificial intelligence can provide feedback, but what effects it has on student performance when it is systematically integrated into an assessment experience.

The technical capacity of language models to generate feedback on specific academic tasks has also been studied. Available results indicate that these systems can recognize relevant elements of a response and produce useful recommendations, but limitations still persist regarding disciplinary accuracy, depth of certain comments, and stability of the generated responses [12]. Therefore, the use of artificial intelligence in evaluation processes must be mediated, so that technology functions as a feedback support and not as an autonomous substitute for teacher judgment. This position is especially interesting when research aims to determine effects on learning, since it allows distinguishing the pedagogical value of feedback from the mere technological availability.

The discussion acquires an additional dimension when evaluation is analyzed from an inclusive perspective. An accessible evaluative process should allow students to understand the instructions, access information about their performance in a timely manner, and have reasonable opportunities to use it to improve their learning. Research on students with disabilities in higher education has shown that generative tools can help overcome certain barriers related to writing, information organization, and comprehension of academic activities; at the same time, they have identified risks linked to errors in responses, access inequalities, and technological dependence [13]. Technology, therefore, does not constitute an inclusive strategy by itself, but rather its contribution depends on the way in which it reduces or reproduces the barriers present in the educational environment.

From the educational design point of view, artificial intelligence has also been suggested as a resource to increase flexibility and provide different modes of interaction with content and instructions [14]. These possibilities are relevant because inclusion in assessment should not be reduced only to accommodations after identifying an individual need. Accessibility must be incorporated from the design of the assessment process, considering the diverse ways in which students receive, interpret, and use information. Previous research on university assessment has pointed out that certain seemingly neutral practices can constitute particular barriers for students with disabilities, which reinforces the need to analyze inclusion as a property of the assessment system and not as an individual characteristic of the student [15].

Despite the growth of research on generative artificial intelligence in higher education, important limitations remain in the available evidence. A relevant proportion of studies focuses on perceptions of usefulness, technological acceptance, quality of generated content, or ad hoc comparisons between human and automated responses. Less frequent are studies that incorporate a structured evaluative intervention and simultaneously compare changes in objective learning outcomes with dimensions related to accessibility, clarity, usefulness, and perceived equity.

In this context, the present study aims to evaluate the effect of formative feedback mediated by artificial intelligence on academic performance and the perception of accessibility and equity of assessment in university students. Specifically, it proposes to compare changes in performance before and after the intervention between students who receive feedback mediated by artificial intelligence and those who receive conventional feedback; to analyze the differences between both groups regarding the clarity, usefulness, accessibility, and perceived equity of the assessment; and to determine whether the effect of the intervention varies according to the initial level of academic performance. It is hypothesized that formative feedback mediated by artificial intelligence will produce a significantly greater improvement in academic performance and a more favorable appraisal of the assessment experience, although the magnitude of this effect may differ according to the students' initial performance.

METHODS

The study was developed under a quantitative approach, with explanatory scope and a quasi-experimental design of non-equivalent groups with pretest–posttest measurement. A total of 128 university students enrolled in the same subject and academic level at a higher education institution in Ecuador participated, distributed into two intact groups of 64 participants. The experimental group received formative feedback mediated by generative artificial intelligence, and the control group maintained conventional feedback provided by the instructor.

Students of legal age, regularly enrolled, who agreed to participate through informed consent and completed the initial and final assessments were included. Those who dropped the course, did not complete a main measurement, or participated in less than 75% of the activities were excluded. The independent variable was the type of feedback. The dependent variables were academic performance, clarity, usefulness, accessibility, and perceived fairness of the assessment. Initial performance was also used as a covariate and moderator.

The intervention lasted eight weeks. During the first week, the pretest was administered; between the second and seventh weeks, six evaluative activities were carried out; and in the eighth week, the posttest and the perception scale were administered. Both groups worked with the same contents, activities, rubrics, criteria, and times. The difference was solely in the feedback modality.

In the experimental group, feedback was generated with the support of artificial intelligence using standardized instructions that incorporated the learning objective, the rubric criteria, and the student's response. The system was to identify strengths, areas for improvement, and concrete recommendations. Each response was reviewed by the instructor before delivery. Artificial intelligence did not assign or modify grades. The control group received feedback prepared directly by the instructor using the same rubric and within equivalent time periods.

Academic performance was evaluated using an objective test of 30 items constructed from the learning outcomes. Equivalent versions were used for pretest and posttest. Five specialists examined the clarity, relevance, and representativeness of the items. Subsequently, a pilot test was conducted and internal consistency was estimated using KR-20, considering values ≥0.70 acceptable.

The evaluative experience was measured using a 20-item Likert scale, with responses from 1 to 5, organized into 4 dimensions: clarity, usefulness, accessibility, and equity. Its construction considered principles of feedback literacy [6,7] and evaluative accessibility [13,15]. Content validity was reviewed by 5 experts, and internal consistency was estimated using Cronbach's alpha and McDonald's omega.

Data were analyzed using descriptive and inferential statistics. Means, standard deviations, medians, frequencies, and percentages were calculated. The effect of the intervention on the posttest was examined using ANCOVA, with group as the factor and pretest as the covariate. Adjusted differences, 95% confidence intervals, p values, and partial eta squared were reported. To determine whether the effect varied according to initial performance, regression with a group × pretest interaction term was used. The low, medium, and high levels were used only for descriptive purposes. α=0.05 was adopted.

The study ensured voluntary participation, confidentiality, and data protection. No personal information was entered into the artificial intelligence system, and all generated feedback was supervised before its academic use.

RESULTS

A total of 128 university students participated, divided into two groups of 64 participants. The experimental group received formative feedback mediated by artificial intelligence, while the control group maintained the conventional teacher feedback procedure. Initial analysis showed practically equivalent mean scores between both groups before the intervention. The experimental group obtained a mean of 59.8 points on the pretest, while the control group reached 60.1 points, with an initial difference of only 0.3 points. This proximity made it possible to consider that both groups started from comparable academic levels in descriptive terms.

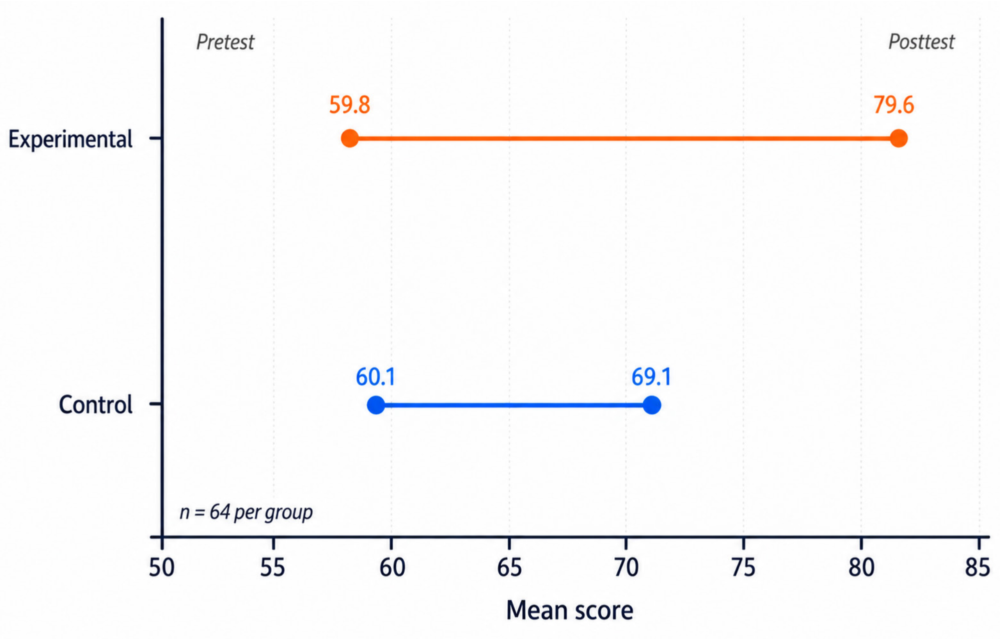
After the intervention, a considerable separation was observed between the trajectories of both groups. The experimental group increased its mean score from 59.8 to 79.6 points, equivalent to a mean gain of 19.8 points. In the control group, the score increased from 60.1 to 69.1 points, corresponding to a mean gain of 9.0 points. Therefore, the improvement observed in the experimental group exceeded that recorded in the control group by 10.8 points. While both groups showed progress with respect to their initial measurement, the change was substantially greater among students who received feedback mediated by artificial intelligence.

The comparison of individual gains allowed us to examine whether the difference between groups depended only on some students with large improvements or whether it corresponded to a more general pattern of performance. The distribution of gain scores showed a shift toward higher values in the experimental group. In this group, most observations were concentrated approximately between 15 and 25 points of improvement, although students with increases greater than 30 points were also identified. In contrast, the control group showed a main concentration around gains lower than 12 points and a greater presence of cases with small changes or close to zero.

The mean gain was 19.8 points in the experimental group and 9.0 points in the control group. In addition to the difference in means, the distributions showed an appreciable separation between the areas of highest density of both groups. This result indicates that the increase associated with the intervention was not limited to a small number of participants, but rather was manifested in a large proportion of the experimental group.

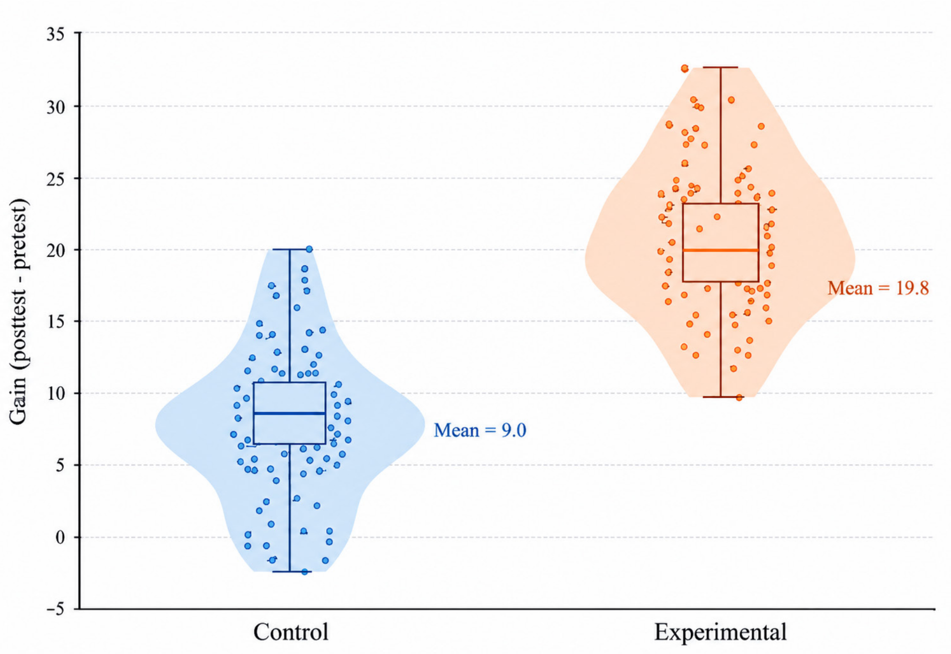
The analysis adjusted using ANCOVA, considering the posttest score as the dependent variable, the group as the factor, and the pretest as the covariate, should be incorporated into the final version with the statistics obtained from the database. The reporting structure will be: F = [value], p = [value], η^2p = [value], 95% CI [lower limit–upper limit]. This analysis will allow establishing whether the observed difference between groups is maintained after statistically controlling for initial academic performance.

Figure 1. Change in mean score between pretest and posttest



Note. The observed means before and after the intervention in both groups are presented. Own elaboration.

Figure 2. Distribution of learning gain



Note. The gain corresponds to the difference between the posttest and pretest scores. Own elaboration.

Student perceptions also showed consistent differences between the two feedback modalities. In the clarity dimension, the experimental group achieved a mean score of 4.47 out of 5, compared to 3.74 in the control group. The difference was 0.73 points. In feedback usefulness, the experimental group recorded its highest score, with a mean of 4.51, while the control group obtained 3.68, resulting in a difference of 0.83 points.

For accessibility, a mean of 4.43 was observed in the experimental group and 3.61 in the control group, with a difference of 0.82 points.

Finally, perceived equity reached a mean of 4.32 in the experimental group and 3.79 in the control group, which represented the smallest difference among the four dimensions, although it maintained a direction favorable to the group that used feedback mediated by artificial intelligence.

Overall, the experimental group presented higher scores on the four dimensions. The widest differences were found in utility and accessibility, whereas the comparatively smaller difference corresponded to equity. This pattern was not concentrated in a single aspect of the evaluative experience but rather encompassed aspects related to understanding the comments, applying them subsequently, accessing information, and perceiving the evaluation process.

Figure 3. Perception of evaluation quality by group.

Note. Mean scores obtained using a Likert scale from 1 to 5. Source: Authors' own elaboration.

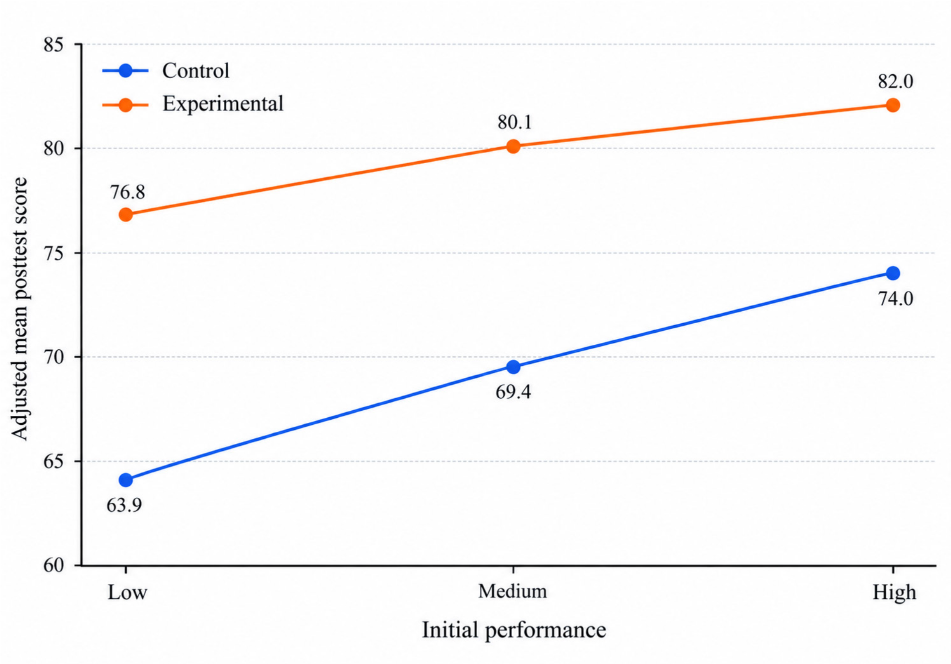
To examine whether the observed effect varied according to initial academic performance, three levels derived from the pretest distribution were analyzed descriptively: low, medium, and high. Among students with low initial performance, the adjusted posttest mean was 76.8 points in the experimental group and 63.9 points in the control group, a difference of 12.9 points. Among students with medium initial performance, the means were 80.1 and 69.4 points, respectively, yielding a difference of 10.7 points. Finally, among those with high initial performance, the experimental group obtained 82.0 points and the control group 74.0 points, resulting in a difference of 8.0 points.

The observed gradient indicated that the difference between the two conditions was greater among students with lower initial performance and progressively decreased as initial performance increased. Although high-performing students also achieved superior results with the intervention, the magnitude of the difference was smaller than that observed at the low and medium initial levels.

The regression model including the interaction between group and initial score should be reported with coefficients derived from the final analysis: β = [value], SE = [value], p = [value], and 95% CI [lower limit–upper limit]. This contrast will formally determine whether the advantage associated with artificial intelligence-mediated feedback changes significantly as a function of initial performance.

In summary, the descriptive results showed three main patterns. First, both groups improved from pretest to posttest, although the gain was greater in the experimental group. Second, feedback mediated by artificial intelligence received higher scores in clarity, usefulness, accessibility, and perceived fairness. Third, the difference between groups was larger among students with lower initial academic performance and progressively decreased among those with higher initial performance.

Figure 4. Interaction between the intervention and initial performance.



Note. Adjusted posttest means are presented according to the descriptive level of initial performance. Authors' own elaboration.

DISCUSSION

The results show that AI-mediated formative feedback was associated with greater improvement in performance than conventional feedback. The experimental group gained 19.8 points from pretest to posttest, while the control group gained 9.0 points. Because initial scores were practically equivalent, the divergence observed during the intervention suggests that the feedback modality may have contributed to the recorded change. This finding is consistent with research showing that AI can produce sufficiently specific and timely comments to support task revision, although it has not demonstrated uniform superiority over human feedback [16].

The distribution of gains also showed that the group difference was not driven solely by a few participants with unusually large improvements. The experimental group exhibited a broader concentration of improvements at higher values, whereas the control group showed smaller gains. This pattern is consistent with reviews that argue that feedback technologies acquire greater value when they are part of successive cycles of execution, comment, and correction [17,18]. In this study, the 6 consecutive activities allowed students to apply the information received in subsequent tasks, which may explain part of the observed advantage.

The results are also consistent with studies linking artificial intelligence, assessment, and learning outcomes. Available evidence indicates that generative tools are more educationally useful when integrated with defined objectives, explicit rubrics, and revision opportunities [19]. Likewise, the speed of feedback can reduce the interval between performance and error correction, although the effectiveness of such feedback depends on the quality of instructions and academic supervision [20]. For this reason, the intervention did not grant grading autonomy to the AI; the instructor reviewed all responses before delivering them.

Student perceptions reinforced the academic results. The experimental group obtained higher scores in clarity, usefulness, accessibility, and equity. The greatest differences appeared in usefulness and accessibility. This result is relevant because the effectiveness of feedback depends on students' understanding of the comments and their ability to subsequently use them to improve, a central principle of feedback literacy [21]. The possibility of receiving structured guidance focused on specific criteria may have facilitated this appropriation.

Regarding accessibility, a difference of 0.82 points was found in favor of the experimental group. The ability of artificial intelligence to expand the possibilities for reformulation, explanation, and adaptation of content has been highlighted as an opportunity to design more flexible educational experiences [22]. However, technology should not be assumed to be inclusive by itself. Accessibility also depends on technological availability, digital competencies, interface quality, and human supervision. In this study, these conditions were partially controlled through a teacher-mediated procedure.

The variation in the difference according to initial performance was among the most relevant results. Among students with low performance, the descriptive advantage of the experimental group reached 12.9 points; at the medium level it was 10.7, and at the high level 8.0. This trend suggests that the intervention could be especially useful for students who begin with greater difficulties. The literature has indicated that the immediate availability of explanations and guidance can promote error resolution and the development of cognitive skills [23,24]. However, this interpretation must be confirmed through the interaction term of the regression model, not merely through descriptive comparison of levels.

In the experimental group, equity was also rated more favorably, although the difference was smaller. This finding may be explained by the fact that evaluative equity depends not only on the clarity of the feedback but also on the consistency of the criteria, opportunities for participation, and assessment conditions. Previous research has shown that apparently homogeneous practices can create different barriers for students [25]. The use of the same rubric and teacher review may have contributed to maintaining common criteria across both conditions.

The main limitations include the use of intact groups, the fact that the study was conducted in a single subject, the relatively short duration, and the use of perceptual measures for accessibility and equity. Furthermore, the results correspond to AI-mediated feedback with teacher supervision, not to a fully automated system, a relevant distinction given the documented differences between human feedback and feedback generated by language models [26].

Overall, the findings support the use of generative artificial intelligence as a support for formative assessment when teacher supervision, explicit criteria, and opportunities to reuse feedback are present. Its main contribution lies not only in improving the academic average but also in being associated with higher levels of usefulness and accessibility, as well as with a possible differential advantage among students with lower initial performance.

CONCLUSIONS

Formative feedback supported by artificial intelligence improved academic performance significantly more than conventional feedback. The experimental group increased from a mean score of 59.8 to 79.6, and the control group from 60.1 to 69.1, with increases of 19.8 and 9.0, respectively.

The intervention also obtained higher scores in the evaluative experience. The experimental group performed better than the control group in clarity (4.47 vs. 3.74), usefulness (4.51 vs. 3.68), accessibility (4.43 vs. 3.61), and perceived fairness (4.32 vs. 3.79), with usefulness and accessibility being the dimensions in which the differences were most evident.

The effect of AI-mediated feedback was not exclusive to high-performing students. Among experimental group students, the largest pre-post score increase was observed in those with low initial performance (12.9 points, compared with 10.7 points for the middle group and 8.0 points for the high group). Thus, AI-mediated feedback may have the potential to improve the performance of students with lower initial performance.

From the perspective of inclusive assessment, artificial intelligence had a more positive effect when incorporated as a support tool in a pedagogical process rather than as an autonomous assessment mechanism. Teacher review, the use of common rubrics, and the regulation of data protection made it possible to maintain pedagogical control while facilitating opportunities for access to feedback.

The results obtained allow us to consider formative feedback mediated by artificial intelligence (AI) as a promising strategy for simultaneously improving learning and some dimensions of the inclusive assessments that must be carried out in higher education. Although these findings are promising, their generalization must be confirmed in future studies that use random assignment of participants, samples from different disciplinary areas, longer follow-up periods, and implementation in populations with differentiated accessibility needs.

REFERENCES

1. Er E, Akçapınar G, Bayazıt A, Noroozi O, Banihashem SK. Assessing student perceptions and use of instructor versus AI-generated feedback. Br J Educ Technol. 2025;56(3):1074-1091. https://doi.org/10.1111/bjet.13558
2. Venter J, Coetzee SA, Schmulian A. Exploring the use of artificial intelligence (AI) in the delivery of effective feedback. Assess Eval High Educ. 2025;50(4):516-536. https://doi.org/10.1080/02602938.2024.2415649
3. Usher M. Generative AI vs. instructor vs. peer assessments: a comparison of grading and feedback in higher education. Assess Eval High Educ. 2025;50(6):912-927. https://doi.org/10.1080/02602938.2025.2487495
4. Zhan Y, Boud D, Dawson P, Yan Z. Generative artificial intelligence as an enabler of student feedback engagement: a framework. High Educ Res Dev. 2025;44(5):1289-1304. https://doi.org/10.1080/07294360.2025.2476513
5. Zhao X, Cox A, Chen X. The use of generative AI by students with disabilities in higher education. Internet High Educ. 2025;66:101014. https://doi.org/10.1016/j.iheduc.2025.101014
6. Qian Y. Pedagogical applications of generative AI in higher education: a systematic review of the field. TechTrends. 2025;69:1105-1120. https://doi.org/10.1007/s11528-025-01100-1
7. Liang J, Stephens JM, Brown GTL. A systematic review of the early impact of artificial intelligence on higher education curriculum, instruction, and assessment. Front Educ. 2025;10:1522841. https://doi.org/10.3389/feduc.2025.1522841
8. Luo J, Zheng C, Yin J, Teo HH. Design and assessment of AI-based learning tools in higher education: a systematic review. Int J Educ Technol High Educ. 2025;22:42. https://doi.org/10.1186/s41239-025-00540-2
9. Daniel K, Msambwa MM, Wen Z. Can generative AI revolutionise academic skills development in higher education? A systematic literature review. Eur J Educ. 2025;60(1):e70036. https://doi.org/10.1111/ejed.70036
10. Deng R, Jiang M, Yu X, Lu Y, Liu S. Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Comput Educ. 2025;227:105224. https://doi.org/10.1016/j.compedu.2024.105224
11. Chen S, Cheung ACK. Effect of generative artificial intelligence on university students' learning outcomes: a systematic review and meta-analysis. Educ Res Rev. 2025;49:100737. https://doi.org/10.1016/j.edurev.2025.100737
12. Drinkwater Gregg K, Ryan O, Katz A, Huerta M, Sajadi S. Expanding possibilities for generative AI in qualitative analysis: fostering student feedback literacy through the application of a feedback quality rubric. J Eng Educ. 2025;114(3):e70024. https://doi.org/10.1002/jee.70024
13. Estévez-Ayres I, Callejo P, Hombrados-Herrera MA, Alario-Hoyos C, Delgado Kloos C. Evaluation of LLM tools for feedback generation in a course on concurrent programming. Int J Artif Intell Educ. 2025;35(2):774-790. https://doi.org/10.1007/s40593-024-00406-0
14. Belkina M, Daniel S, Nikolic S, Haque R, Lyden S, Neal P, et al. Implementing generative AI (GenAI) in higher education: a systematic review of case studies. Comput Educ Artif Intell. 2025;8:100407. https://doi.org/10.1016/j.caeai.2025.100407
15. Dai W, Tsai YS, Lin J, Aldino A, Jin H, Li T, et al. Assessing the proficiency of large language models in automatic feedback generation: an evaluation study. Comput Educ Artif Intell. 2024;7:100299. https://doi.org/10.1016/j.caeai.2024.100299
16. Steiss J, Tate T, Graham S, Cruz J, Hebert M, Wang J, et al. Comparing the quality of human and ChatGPT feedback of students' writing. Learn Instr. 2024;91:101894. https://doi.org/10.1016/j.learninstruc.2024.101894
17. Sembey R, Hoda R, Grundy J. Emerging technologies in higher education assessment and feedback practices: a systematic literature review. J Syst Softw. 2024;211:111988. https://doi.org/10.1016/j.jss.2024.111988
18. Xia Q, Weng X, Ouyang F, Lin TJ, Chiu TKF. A scoping review on how generative artificial intelligence transforms assessment in higher education. Int J Educ Technol High Educ. 2024;21:40. https://doi.org/10.1186/s41239-024-00468-z
19. Weng X, Xia Q, Gu M, Rajaram K, Chiu TKF. Assessment and learning outcomes for generative AI in higher education: a scoping review on current research status and trends. Australas J Educ Technol. 2024;40(6):37-55. https://doi.org/10.14742/ajet.9540
20. Lee SS, Moore RL. Harnessing generative AI (GenAI) for automated feedback in higher education: a systematic review. Online Learn. 2024;28(3):82-106. https://doi.org/10.24059/olj.v28i3.4593
21. Dawson P, Yan Z, Lipnevich A, Tai J, Boud D, Mahoney P. Measuring what learners do in feedback: the Feedback Literacy Behaviour Scale. Assess Eval High Educ. 2024;49(3):348-362. https://doi.org/10.1080/02602938.2023.2240983
22. Stefaniak JE, Moore SL. The use of generative AI to support inclusivity and design deliberation for online instruction. Online Learn. 2024;28(3):181-206. https://doi.org/10.24059/olj.v28i3.4458
23. Essel HB, Vlachopoulos D, Essuman AB, Amankwa JO. ChatGPT effects on cognitive skills of undergraduate students: receiving instant responses from AI-based conversational large language models (LLMs). Comput Educ Artif Intell. 2024;6:100198. https://doi.org/10.1016/j.caeai.2023.100198
24. Escalante J, Pack A, Barrett A. AI-generated feedback on writing: insights into efficacy and ENL student preference. Int J Educ Technol High Educ. 2023;20:57. https://doi.org/10.1186/s41239-023-00425-2
25. Tai J, Mahoney P, Ajjawi R, Bearman M, Dargusch J, Dracup M, et al. How are examinations inclusive for students with disabilities in higher education? A sociomaterial analysis. Assess Eval High Educ. 2023;48(3):390-402. https://doi.org/10.1080/02602938.2022.2077910
26. Carless D, Boud D. The development of student feedback literacy: enabling uptake of feedback. Assess Eval High Educ. 2018;43(8):1315-1325. https://doi.org/10.1080/02602938.2018.1463354

FUNDING

None.

CONFLICT OF INTEREST

There are none.

AUTHORSHIP CONTRIBUTIONS

Conceptualization: Darwin Joel Lozada López.
Research: Darwin Joel Lozada López.
Methodology: Darwin Joel Lozada López.
Writing – original draft: Darwin Joel Lozada López.
Writing – review and editing: Darwin Joel Lozada López.