ISSN: 2458-8989

Natural and Engineering Sciences

NESciences, 2026, 11 (3): 303-315 doi: 10.28978/nesciences.263026

Evaluating an Artificial Intelligence–Assisted Mathematical Reasoning Framework for Biological Systems Modeling and Engineering Education

Reinaldo Antonio Guerrero-Chirinos[1*], Hugo Alfredo Pérez-Benítez [2], Kenneth David Suarez-Rivera[3], Jean Pierre Reyes-Carrión [4], Abraham Coila-Torres[5], Mario Aguilar-Fruna[6]

[1] Universidad Técnica Particular de Loja, UTPL, Ecuador.
E-mail: raguerrero12@utpl.edu.ec

[2] Universidad de Guayaquil; Universidad Bolivariana del Ecuador, Ecuador.
E-mail: hugo.perezb@ug.edu.ec / haperezb@ube.edu.ec

[3] Universidad Técnica Particular de Loja, UTPL, Ecuador. E-mail: kdsuarez1@utpl.edu.ec

[4] Universidad Nacional de Educación, UNAE, Ecuador. E-mail: jean.reyes@unae.edu.ec

[5] Universidad Nacional de Juliaca, Peru. E-mail: acoilat.doc@unaj.edu.pe

[6]Escuela de Educación Superior Pedagógica Pública Juliaca, Peru.
E-mail: aguilarfrunamario@gmail.com

Abstract

This study examines an artificial intelligence (AI)-supported mathematical reasoning model for biological systems modeling and engineering education. In particular, it examines the effects of a PBL framework on undergraduate engineering students solving modeling problems that also include the use of an algorithm designed to aid mathematical reasoning. The study was conducted using a quasi-experimental design featuring a control group (n = 35) and an experimental group (n = 36) drawn from 71 students in the Mathematical Foundations course at Universidad Técnica Particular de Loja. Assenting to group equivalence at baseline (Levene's test; Student's t-test: $t = 0.425$; $p = 0.672$), both groups participated in an academic term using an AI-mediated PBL approach. Positive post-intervention test results reveal a statistical advantage for the experimental trials ($M = 8.5$) over the control trials ($M = 7.4$) in statement interpretation ($t = -2.295$; $df = 69$; $p = 0.012$) and in algebraic modeling ($t = -2.360$; $df = 69$; $p = 0.015$). Graphical representation also improved ($p = 0.042$), but it remains the most challenging area. The results highlight the value of AI as a ‘cognitive scaffold’ that does not replace teacher mediation but complements it to enhance learning, and they have implications for how curricular content is governed to foster digital literacy and ensure proper use of AI for learning in the engineering domain.

Keyword
Problem-Based Learning; artificial intelligence; mathematical reasoning; linear inequalities; engineering education; quasi-experiment

Available Online: 01/09/2026

**Corresponding Author: Reinaldo Antonio Guerrero-Chirinos, E-mail: raguerrero12@utpl.edu.ec*

Introduction

Artificial Intelligence (AI) has been rapidly transforming engineering education, especially in modeling complex biological systems. The applications of AI in mathematical reasoning hold promise for enhancing students' understanding and problem-solving skills within mathematical frameworks. This study examines how an AI-assisted approach can help undergraduate engineering students develop essential modeling skills. There is hardly anything more disruptive in the university classroom than the arrival of generative artificial intelligence. A set of tools designed over less than two years to provide conversational, step-by-step solutions to equations, or explanations customized to a listener's proficiency, has become such a regular part of the academic world for thousands of students that they are hardly seen as a novelty. This rate of adoption has, in many instances, proven to be faster than official institutions' ability to consider the pedagogical implications of this, particularly in cases where the tools are not new but simply added onto existing pedagogically established approaches of instructional delivery, such as hands-on learning or active learning, and are superimposed upon them. In many situations, the adoption rate is faster than the ability of official institutions to reflect on the pedagogical implications, so they are not making a revolution by entirely replacing pedagogically groundbreaking approaches to instructional delivery, but rather are superimposed on the established methodologies. This can be an important issue to consider in Problem Based Learning (PBL).

This methodology has been used for more than four decades, and its pedagogical value is justified by the active management of students' reasoning rather than by mere replication (Ordóñez Fernández et al., 2026; Tapia-Vélez et al., 2020). The challenges it has encountered have been persistent in disciplines that rely heavily on quantitative methods, where implementation is hindered by curricular rigidity and a lack of articulation between theory and practice (Ordóñez Fernández et al., 2026; Terán Ortiz, 2021). What's not in question in this study, in fact, is whether PBL works, that is, whether it's reasonably effective as it is; instead, the question is whether using this PBL in conjunction with an AI genie accentuates its impact levels, or whether it adds a dependence on the genie that slowly undermines the cognitive intent of PBL.

The study of linear inequalities can be pursued at the university level and provides a useful context for this discussion. The literature presents a consensus that students' problems are not about using mechanical techniques to manipulate algebraic equations but rather about a poor conceptual understanding of the properties of inequalities and feasible regions (Rodríguez & Ravelo, 2026; Terán Ortiz, 2021). This distance between "scholarly knowledge" and "school knowledge", as Fuentes González (2017) and Arévalo and Rojas (2021) put it, results in systematic errors that do not disappear with repeated practice in the exercise routine. Some attempts at a remedy have been made with virtual laboratories and digital modules (Ayala Bendezu, 2020; Hernandez Hurtado, 2014), but few are sufficiently systematic to establish causality for the intervention method itself, that is, through an experimental design.

New foundation language models have been added, including GPT-4 and its variants (Mora-Delgado et al., 2026). This is because, unlike dynamic geometry software that provides feedback based on fixed definitions, a generative AI assistant can sustain an ongoing dialogue, rephrase the question if some information is not understood, and also provide alternative explanations if the initial one is not working out. It is essential to set up an ethical framework in relation to the accuracy of the options that can be generated by these systems, the privacy of the students' information, and the risk that the students, at the very least, can pass off to the system the very thinking power they would use with their own hands. However, AI, when designed and used correctly, does not replace critical thinking but rather enhances it (Pujol et al., 2026; Szczesniewski et al., 2024), and, if implemented without judgment, it could have the opposite effect. A less-studied aspect of educational technology is the influence of executive functions and study strategies on knowledge in Mathematical Problem Solving (MPS; Martín-Requejo et al., 2023).

The pedagogical structure on which any innovative approach, including AI, depends hinges on teachers' ability, as trainers and actors in learning, to prepare to use it effectively (Porlán et al., 2010). Despite this evidence, there remains a clear gap in the literature: PBL and generative AI are mostly studied separately, and few studies that examine them together employ a quasi-experimental design that allows for identifying all factors other than PBL/generative AI that may affect the results (Trullás et al., 2022; Robayo Avendaño, 2026). The majority of previous studies are either scoping reviews or case studies without comparison, which means this study aims to provide a better answer to this question than most previous studies. In this context, the current study aimed to examine the impact of using the model of

problem-based learning by means of artificial intelligence on the students' learning results and mathematical reasoning of UTPL engineering students taught linear inequalities, compared to the students taught traditionally.
The research question used in a PBL classroom can be formulated as follows: How can teachers' use of AI tools in a PBL classroom affect academic achievement and mathematical reasoning skills for solving linear inequalities? Based on the working hypothesis that a digitalized model built around collaborative problem-solving with real-time algorithmic feedback would lead to better summative performance measures and greater intellectual autonomy that generalizes to other mathematical content, learners were divided into three experimental groups. The paper is organized as follows: It outlines the conceptual framework that connects PBL with MathWS and AI-mediated PBL, the quasi-experimental methodological design, and the ethics involved in implementing this design. It then presents the pre-test and post-test results, the analysis by reasoning dimension, and the correlation between reasons for using AI tools and achieved performance, stratified by frequency of use at school. Finally, the study concludes with a discussion that contextualizes the findings within the literature review and draws conclusions directed toward the institution's curricular governance.

2.Conceptual Framework

If one takes a step to understand the cause of a student's frequent trouble with a given linear inequality, one will find it difficult to resist a murky assumption, namely that problems involving algebra are not always arithmetic. The didactic situation theory was first developed by Brousseau and later adopted in higher education by Moreno-Sánchez (2011), who stated that mathematical knowledge aggregates during a situation that requires knowledge for its solution. Therefore, it does not matter whether PBL is just one of many instructional techniques; rather, PBL is the educational translation of an epistemological principle, as teachers learn and grow better when addressing problems than when memorizing processes (Ordóñez Fernández et al., 2026; Tapia-Vélez et al., 2020).
PBL has an additional dimension when applied to engineering. A student programming for resource optimization, modeling structural stability, or assessing technical feasibility using systems of inequalities is not merely performing algorithms; they are explaining the reasons behind their choice of one algorithm over another, which is exactly what the discipline requires of its future practitioners (Sepulveda et al., 2021). In contrast, empirical experiences with linear inequalities suggest a cyclic process: students' logical struggles do not appear to stem from a lack of logical ability but from an incomplete understanding of the formal properties of inequalities, the concept of a feasible region, and interval notation (Rodríguez & Ravelo, 2026; Terán Ortiz, 2021). To make sense of this phenomenon, Arévalo and Rojas (2021) propose the Mathematical Working Space (MWS) concept, which they describe as a "personal construction formed through the interaction of the student with physical or digital artifacts, mediated and intentional by the instructor."
From this perspective, a technological instrument cannot serve as a shortcut that bypasses the transition from school knowledge to scientific or scholarly knowledge. Rather, it functions as scaffolding (Vygotsky's sense of the term), facilitating the transition from school to scientific knowledge (Fuentes González, 2017). Generative AI adds a new qualitative dimension to previous digital tools. A conversational assistant can provide feedback tailored to a student's error, rather than the same feedback for all students, and allow an inequality to take shape through the student's trial, question, and reformulation (Chávez-Mostajo, 2025; Robayo Avendaño, 2026). Large Language Models are very effective information synthesizers, according to Mora-Delgado et al. (2026), but they can be detrimental to learning if teachers do not ensure that students do not use them as a "black box." In contrast, the learning analytics usually found in such systems enable teachers to uncover patterns of errors that are not easily noticeable in the traditional classroom (Robayo Avendaño, 2026; Uzun et al., 2025). Third, the educational-technology literature underdevelops the conceptual axis of the importance placed on neurocognitive factors.
In the context of mathematical problem-solving, Martín-Requejo et al. (2023) demonstrated that both low- and higher-level cognitive functions, such as Executive Functions, Emotional Intelligence, and Study Habits, are robust predictors of performance among science students. AI could make a more explicit and distinct contribution from a particular perspective rather than "teaching content," specifically by alleviating cognitive load from repetitive, procedural activities, thereby freeing mental resources that would otherwise be dedicated to metacognition and strategic analysis of a problem (Martín-Ruiz & González-Valenzuela, 2024). This is an important change in the engineering field,

where choices must be based on careful judgment and evidence grounded in mathematics. Unfortunately, none of these benefits are obtained automatically when an AI tool is added to the classroom. Alvarado Martínez et al. (2026) argue that the introduction of the concept of assessment of learning into the classroom should not be reduced to an average-oriented exercise, but rather treated as a subject matter that develops proper "statistical literacy" for critically reading real-world circumstances. If this argument holds for this study, the conclusions regarding the impact of AI are valid only for a methodological design that successfully reduces and controls confounding elements, which is ideal for quasi-experimental designs because of their ability to compare teaching methods with reasonable controls (Caballero-Cobos & Llorent, 2022).

Automating education can also create algorithmic bias if it is designed without the principles of equity and transparency, as noted by Marcano-Millán et al. (2023). A lack of faculty trained to use these tools can hinder innovation, preventing it from accelerating learning and instead making it a distraction, as Pujol et al. (2026) emphasized. In this instance, the role of the educator in this type of reading PBL is to design learning experiences in which AI should not "solve" tasks for students but should support them in understanding the logic of solving linear inequality problems. In conclusion, the glue that holds all four components of the presented research together is that deep learning of linear inequalities is not simply the outcome of algorithmic practice; it is the result of the interplay among student curiosity, problem structure, and the quality of the technological feedback they receive. The concept tested by the methodological design below is as follows.

3. Materials and Methods

The method used was a cross-sectional quasi-experimental design, comparing learning with Problem-Based Learning (PBL) using artificial intelligence to conventional learning. Two classes of undergraduate engineering students were involved, and each class was randomly assigned to one of the experimental conditions. Students' mathematical reasoning was the focus of data collection, with an emphasis on modeling biological systems. For several reasons related to the university setting, including the fact that students were already in academic sections, a quasi-experimental design was used instead of a purely experimental design. The sections could not be randomly assigned to the various conditions individually, but it was possible to randomly assign the sections to the conditions.

3.1 Population and Sample

The reference population consisted of students enrolled in the Mathematical Foundations course for Engineering Degree programs at Universidad Técnica Particular de Loja (UTPL). Because this was a census of all students enrolled in the available sections of the class during the school year, the population size had to be adjusted for a finite population, and a representative sample of 71 students was obtained. Using simple random sampling at the section level, which provides a reasonable degree of equity in the input variables and less voluntary selection bias, the students were randomly assigned to an experimental group ($n = 36$) and a control group ($n = 35$).

3.2 Data-Collection Techniques and Instruments

The data were obtained from an academic performance test designed to measure competency indicators related to linear inequalities and administered twice (before and after the intervention) within a systematic questioning structure. The instrument was reviewed by experts in mathematics didactics and university pedagogy to ensure its alignment with the aims of the degree program. The instrument was tested for psychometric rigor using Cronbach's alpha to assess item consistency and to confirm that the items were an appropriate measure of the mathematical-reasoning construct. It was also assumed that between-group comparisons were valid, using Levene's test to verify homogeneity of variances between the control and experimental groups, which was considered a criterion for the valid use of robust parametric tests such as the Student's t-test.

3.3 Intervention Procedure

The intervention spanned 1 full school term. This experimental group was taught using a PBL model consisting of four phases: needs identification, collaborative exploration, hypothesis generation, and metacognitive reflection, throughout which a generative AI assistant was used as an iterative support tool. The assistant provided instant feedback on attempted solutions and, when the group needed extra practice, generated additional problems related to those in the given text. It also assisted in translating

English sentences into algebraic expressions. The control group, on the other hand, was instructed on the same thematic unit, except that the algorithmic-mediation element was omitted from the course's learning method. For both groups, the number of instructional hours and the content covered in the course syllabus were kept constant, so that the teaching strategy used was the variable to be manipulated, as far as possible. The data were analyzed using SPSS version 31.

The analysis plan consisted of two phases. The descriptive stage involved determining summary scores, means, standard deviations, and frequency measures, and thus characterizing the initial and final levels of students in each group. The second, inferential stage involved the use of the Student's t-test for independent samples, which compared the mean scores of the control and experimental groups in both pretest and posttest. The statistical decision criterion was $p < .05$, meaning that the null hypothesis, H0: No Difference, would be rejected if the results indicated an advantage of the AI method of mediation over the traditional method. Additionally, a Pearson correlation coefficient was computed to explore the dose–response effect of the intervention on the relationship between the reported frequency of using an AI tool and the posttest results.

3.5 Ethical Considerations

The execution of this research with engineering students at University Técnica Particular de Loja (UTPL) prioritized the respect and integrity of those involved, guided by ethical principles. Clear information about the study's purpose, the voluntary nature of participation, and the right to withdraw without any consequences for a student's course evaluation was provided to students before their involvement in the study, in line with informed-consent principles found in the literature on education-research ethics, which are treated as non-negotiable (Inglada Galiana et al., 2024). To ensure broad adherence to the guidelines outlined by Marcano-Millán et al. (2023) regarding data protection within algorithm-driven educational environments, individual performance data were anonymized and kept confidential. Regarding the use of a generative AI tool, a responsible-use protocol was developed that restricts the tool to pedagogical support and the production of practice problems, consistent with the cautions of Szczesniewski et al. (2024) regarding the ethical boundaries of case simulation with AI. The findings of Pujol et al. (2026) on the risk of the digital divide becoming a new form of educational inequality were taken into account, with an effort to implement it equitably for participants, providing access to the technological tool under the same conditions, regardless of their social or socioeconomic background or previous personal experience. The goal of these ethical considerations was to base the design of effective and equitable PBL and AI training methods on the social duty that university educational research is meant to serve.

4. Results

This section first compares the control and experimental groups before the intervention (pretest), then analyzes how effectively the intervention promotes success (posttest), then dissects the intervention by achievement level and by specific mathematical reasoning skills, and finally examines the relationship between the frequency of application of the AI tool and resulting mathematical success.

4.1 Baseline Comparison (Pretest)

A preliminary diagnostic evaluation determined that both groups began at a statistically identical baseline level of prior knowledge. An independent Student's t-test was conducted to compare the means of the control and experimental groups before the intervention, yielding $t = 0.425$ ($df = 68$; $p = 0.672$), which was not statistically significant. Homogeneity of variances between the two groups was also confirmed by Levene's test ($F = 0.474$, $p = 0.493$); parametric statistics were then used.

Table 1. Pretest Results: Levene's and Student's t-Tests Between the Control and Experimental Groups

Test	F / t	df	Sig. (p)	Control Group M	Experimental Group M
Levene	0.474	—	0.493	5.00	4.86
Student's t	0.425	68	0.672	5.00	4.86

Note. Baseline performance comparison (pretest). N = 71. Homogeneity of variances confirmed via Levene's test ($p > .05$).

Beyond its statistical basis, the overall pretest average (below 5 points in both groups) is informative. The Ecuadorian grading scale is 1 to 10, with the minimum passing grade being 7, suggesting that an average close to 5 for most students may indicate they lacked the required conceptual foundation expected from a university-level mathematical foundations unit. The findings align with those of others who have discussed persistent struggles in solving quadratic and linear systems (e.g., Rodríguez and Ravelo, 2026) and point to broader issues in the transition from secondary to postsecondary education. Beyond the overall error pattern 'fewer than marks,' an analysis of the errors made in the pretest indicates that there are at least four common patterns of difficulty: literal interpretation of problem statements; converting contexts into algebraic representations; not knowing the conventions of the graphically presented inequality; and having no criteria for determining whether a region of the plane is a valid solution to the system. These four factors of error are not unrelated; a student who cannot formulate a translation would certainly have a problem solving a literal equation graphically, and a sequence of pedagogical interventions would need to address these errors in sequence rather than as isolated errors.

4.2 Effect of the Intervention (Posttest)

This all changed following the intervention. The experimental group averaged 8.5 points on the post test while the control group averaged 7.4 points (Table 2). However, the difference was found to be statistically significant, using a Student's t-test for independent samples (t = -2.295; df = 69; p = 0.012), and the difference is within a 95% confidence interval of -2.056 to -0.144, which excludes the null value of zero, further supporting the conclusion that the difference is not random.

Table 2. Posttest Results: Levene's and Student's t-Tests Between the Control and Experimental Groups

Test	F / t	df	Sig. (p)	95% CI Lower	95% CI Upper	Control M	Experimental M
Levene	—	—	—	—	—	7.40	8.50
Student's t	-2.295	69	0.012	-2.056	-0.144	7.40	8.50

Note. Final performance comparison (posttest). Statistical significance: $p = .012$ ($p < .05$). 95% CI of the difference: [-2.056, -0.144].

One effect that the aggregate figures might mask should be noted: In contrast to their own pre-test scores, the control group's average increased significantly from 5.00 to 7.40 points; the 2.4-point improvement was achieved exclusively through the lecture method. This eliminates any interpretation that the traditional approach is ineffective and even brought the control group's average above the minimum passing grade. The experimental group did not fail the conventional teaching; rather, instruction mediated by AI showed an additional increase of about 1.1 points, on top of the already significant improvement, over a roughly 20-point scale from 1 to 10, for a group of students who were beginning their learning process at a level comparable to that of the control group.

This additional 1.1 points on a 10-point scale, relative to a group of students who, not long ago, began their learning at a level comparable to the control condition, is far from trivial pedagogically. This enhancement was clearly and specifically focused on students' capacity to understand problems in a real-world context, translate such problems into inequalities, graphically depict the constraints at play, and determine solution regions based on those constraints with reasoned judgment. These results align with existing studies that firmly established that PBL holds great potential to promote logical reasoning and genuine learning in Mathematics (Faicán, 2022; Gómez & Espinoza, 2022), implying that the AI aspect is not seen as a substitute for the logical reasoning aspect and is integrated where necessary in the PBL process to facilitate cognitive flow.

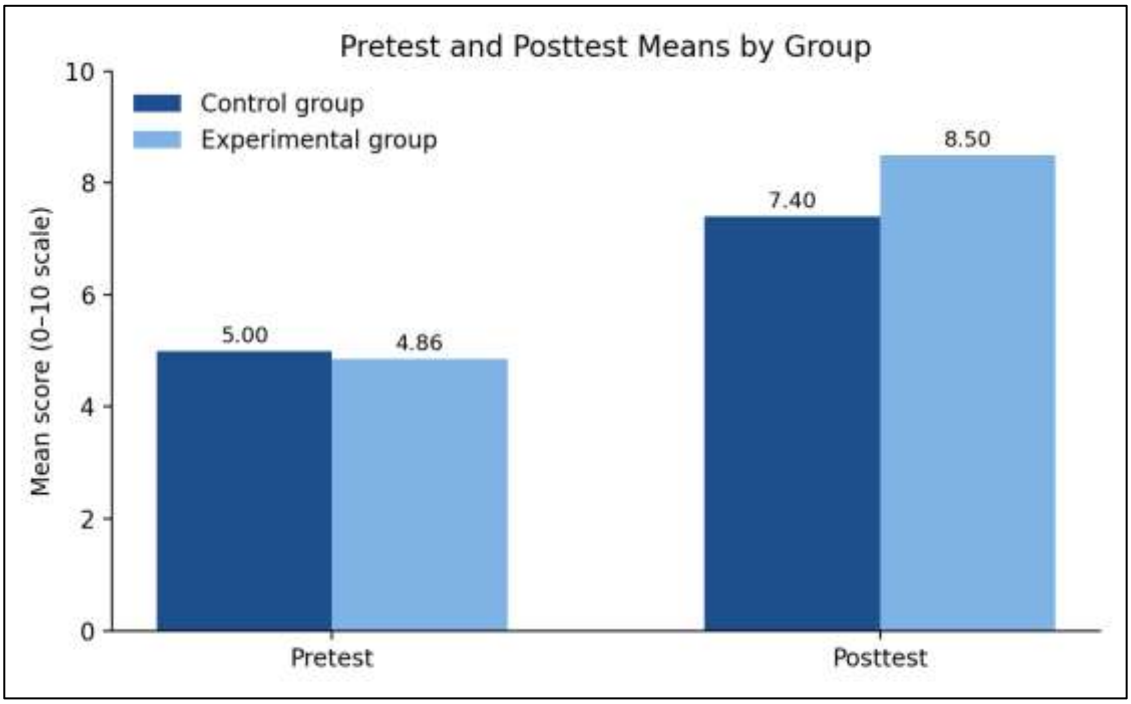

Figure 1. Comparison of Performance Means (Pretest and Posttest) Between the Control and Experimental Groups

4.3 Achievement Levels

The changes in the proportion of students by achievement level (Table 3) provide a useful complement to the analysis of means. In the pre-test, the proportion of students who received an "insufficient" level in both groups exceeded half, and the proportion receiving an "outstanding" level was low in both groups, with only 2.8% and 2.9% receiving it, respectively. The experimental group experienced a significant change after the intervention: the "insufficient" level fell from 19.0% to 8.3%, the "good" level rose from 11.1% to 52.8%, and the "outstanding" level increased almost eightfold to 22.2%. The control group also made progress (the "insufficient" level fell to 17.1%, and the "good" level increased to 40.0%), but the gains were less pronounced than in the group receiving support.

Table 3. Comparison of Achievement Levels, Pretest Versus Posttest, by Group

Achievement Level	Pretest (Exp.)	Posttest (Exp.)	Pretest (Control)	Posttest (Control)
Insufficient (< 5)	**52.8%**	**8.3%**	**51.4%**	**17.1%**
Regular (5–6.9)	**33.3%**	**16.7%**	**34.3%**	**34.3%**
Good (7–8.9)	**11.1%**	**52.8%**	**11.4%**	**40.0%**
Outstanding (≥ 9)	**2.8%**	**22.2%**	**2.9%**	**8.6%**

Note. Substantial improvement in the "Good" and "Outstanding" levels for the experimental group following AI mediation.

The redesign of this pattern and its ability to show the percentage of students shifting from academic-risk levels to solid mastery are, in some ways, more informative than a single comparison of means. It not only reveals that the improvement did not restrict itself to a small number of very high-performing students but also shows that it moved a broad swath of students off academic-risk and into solid mastery. In other words, the program seems to have had a different impact on students with the lowest initial performance than on students in other groups, potentially reducing gaps in school performance within a year, a finding that should be tracked in future replications of this study.

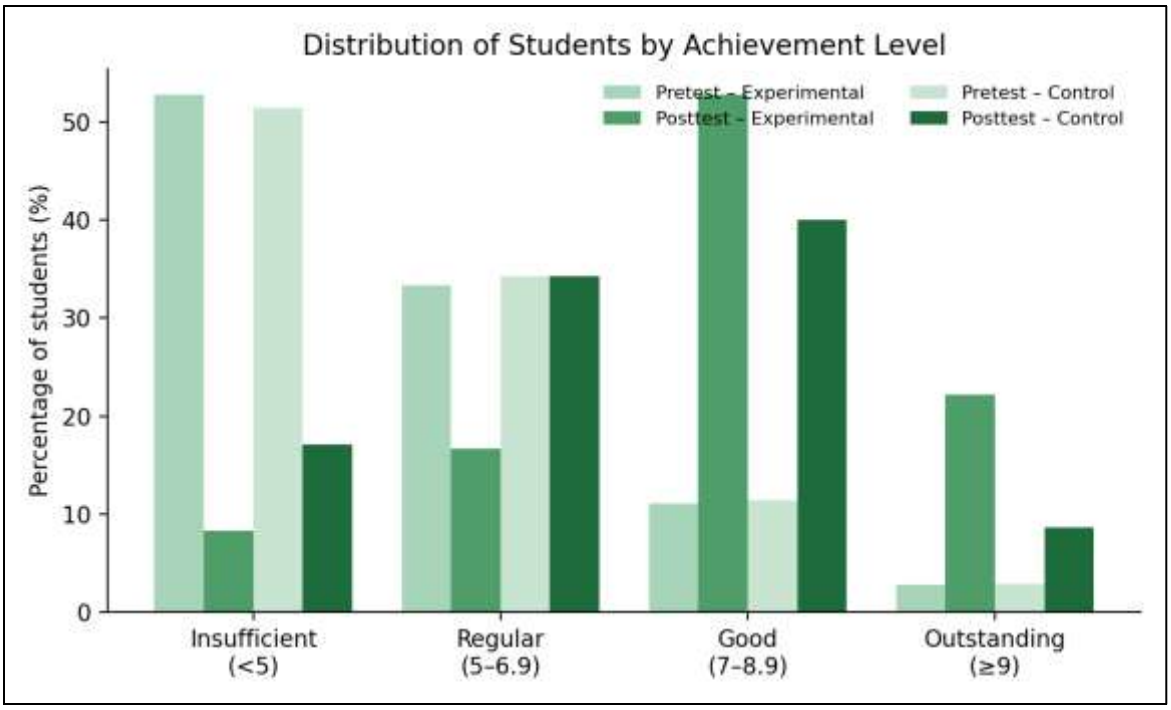

Figure 2. Percentage Distribution of Students by Achievement Level, Pretest Versus Posttest, Experimental Versus Control Group

4.4 Performance by Dimension of Mathematical Reasoning

An analysis of performance within each dimension of mathematical reasoning (Table 4) pinpoints the impact each dimension has on overall performance. Three sub-competencies are measured: statement interpretation, algebraic modeling, and graphical representation. The first two showed a significant and large difference between the experimental group's mean scores and the control group's mean scores, both of which exceeded the 0.5 criterion for statistical significance, with statistical significance being accepted for statement interpretation ($M = 8.8$ versus 7.5, difference = 1.3 and $p = 0.008$) and algebraic modeling ($M = 8.4$ versus 7.2 difference = 1.2 and $p = 0.015$). The difference in graphical representation, however, was also significant ($M = 8.3$ versus $M = 7.5$; difference = 0.8; $p = 0.042$), but it was less noticeable than for the other two dimensions.

Table 4. Performance by Dimension of Mathematical Reasoning (Posttest)

Dimension	Experimental M	Control M	Difference	Sig. (p)
Statement interpretation	8.8	7.5	1.3	0.008
Algebraic modeling	8.4	7.2	1.2	0.015
Graphical representation	8.3	7.5	0.8	0.042

Note. AI appears to function primarily as scaffolding for translating natural language into algebraic language, with a more modest effect on graphical representation.

It is not a coincidence that this is asymmetrical in its dimensions and should be read carefully. By nature, a conversational AI assistant should be better at comprehension tasks or symbolic manipulation (in essence, language processing) than at developing spatial intuition for graphically portraying a feasible region. This might account for the gains in statement interpretation and algebraic modeling being nearly double those for graphical representations. The persistence of this gap implies that Generative AI, applied in this intervention, does not stand in for the need for dynamic-geometry programs like GeoGebra or Desmos, but rather supports them as part of the whole, fitting with the perspectives on the consolidation of concepts that require three-dimensional visualization presented by Gómez and Espinoza (2022) regarding the relevance of such programs in engineering problems, and, as they explain, how Generative Artificial Intelligence fits into that environment is complementary.

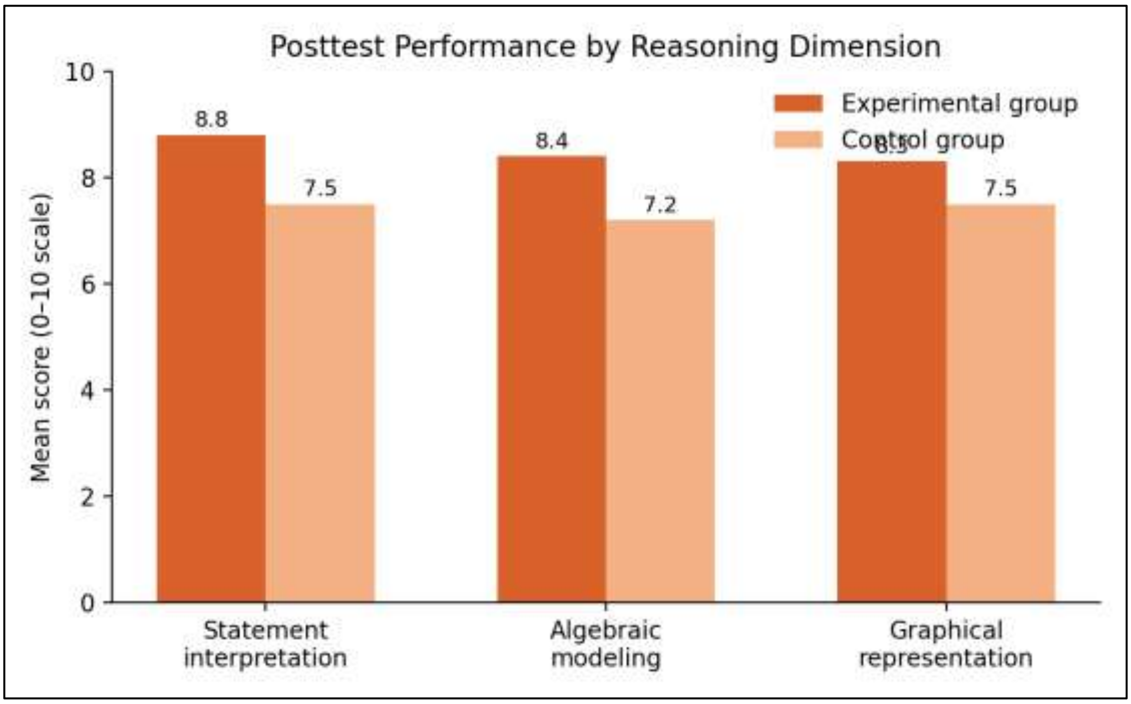

Figure 3. Comparative Performance by Dimension of Mathematical Reasoning (Posttest)

4.5 Correlation Between AI Use, Performance, and Autonomy

Finally, the study examined the correlations between self-reported frequency of AI tool use and posttest scores, as well as between self-reported frequency of AI tool use and students' perceived level of autonomy (Table 5). Hours of AI use was highly correlated with performance ($r = 0.642$, $p = 0.001$) and moderately correlated with autonomy ($r = 0.589$, $p = 0.003$).

Table 5. Correlation Between Frequency of AI Use and Student Performance and Autonomy

Variable	Correlation (r)	Sig. (p)	Interpretation
Hours of AI use (vs. performance)	0.642	0.001	Moderate positive correlation
Hours of AI use (vs. autonomy)	0.589	0.003	Moderate positive correlation

Note. Frequency of AI-tool use is positively associated with achievement of learning objectives and with perceived student autonomy.

The moderate correlation observed in this study, rather than a stronger one, is actually the most reassuring finding. If AI-mediated PBL were seen as efforts to teach learners to substitute cognitive effort, a correlation near 1:1 would have suggested a mechanical dependency between tool use and performance. The moderate magnitude, more similar to what Robayo Avendaño (2026) observes regarding other learning analytics in PBL settings, also aligns more closely with the “dosage” model: regular, but not excessive, use of AI produces better results and does not simply mean that “the more, the better” for learning.

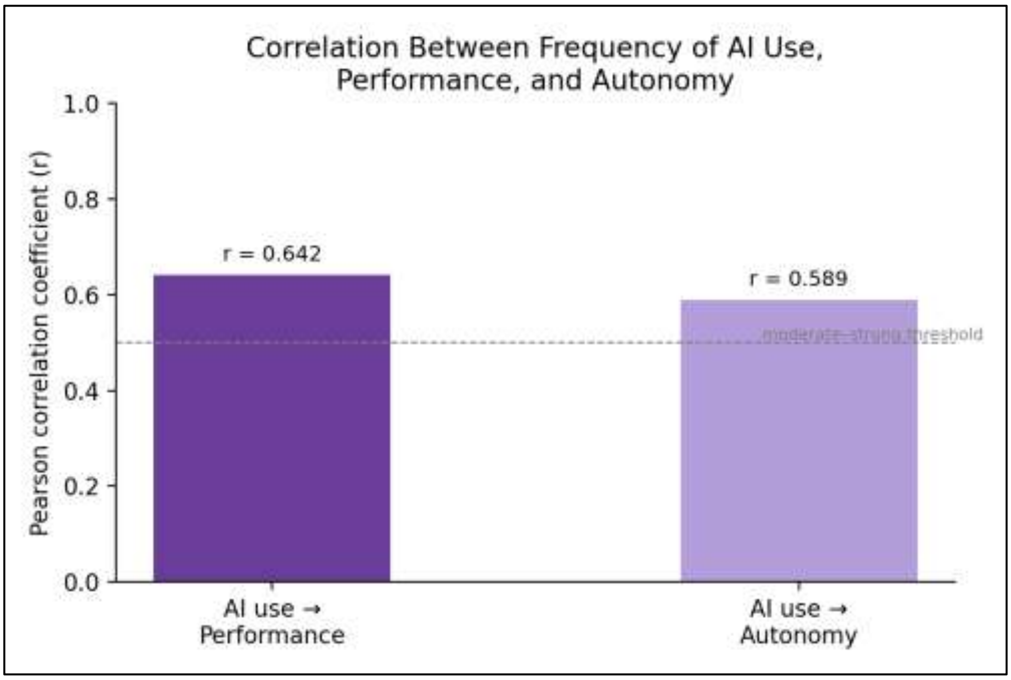

Figure 4. Correlation Between Frequency of AI Use, Posttest Performance, and Perceived Autonomy

A complementary way to assess the size of the intervention is to examine the pretest-to-posttest growth for each group (Figure 5). The control group showed a statistically similar 2.4-point improvement with the conventional approach to instruction, while the experimental group gained 3.64 points with the AI approach to PBL — a gain that, although smaller, is approximately 1.5 times greater than the improvement in the control instruction group.

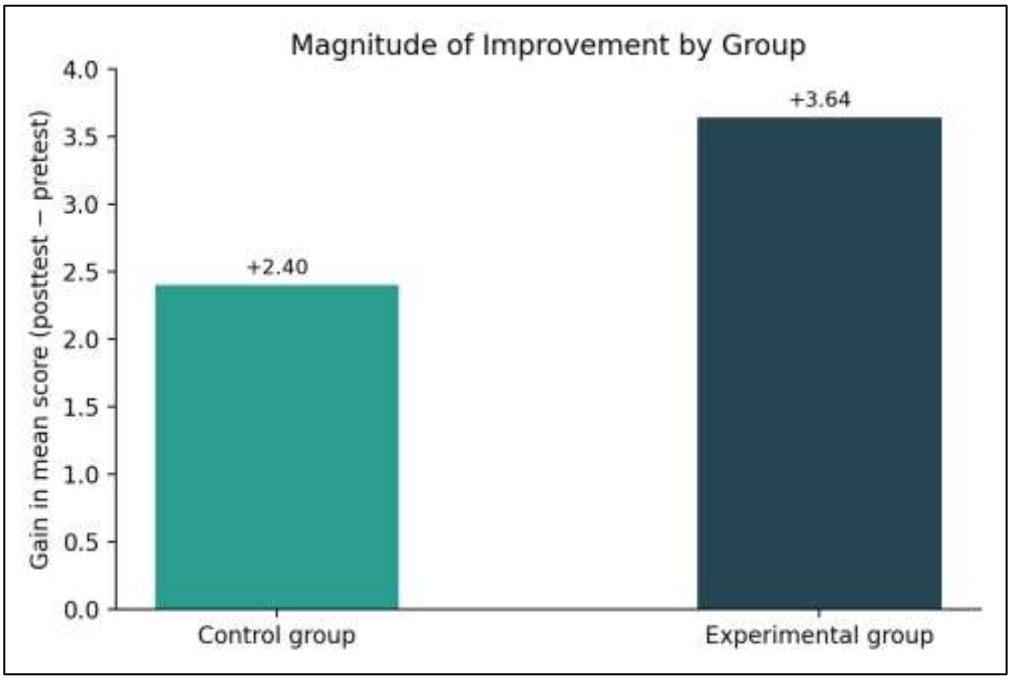

Figure 5. Magnitude of Pretest-to-Posttest Improvement by Group

5.Discussion

The findings of this research broadly support the proposed hypothesis that an AI-mediated PBL approach yields a significant performance improvement in linear inequalities compared with traditional instruction. However, a careful examination of these data raises discussion of at least three paths that resonate with the literature reviewed. The effect size (1.1 points on a 10-point scale) is statistically significant but not large enough to conclude that AI replaces teachers, as proposed by Mora-Delgado et al. (2026). Rather, it acts as a cognitive catalyst and can never replace teachers. The effect was large, but it was measured in an ecological setting—a real classroom, a common instructor for both conditions, and similar, specific curriculum content—in contrast to studies reporting gains under highly controlled conditions. Second, because the improvement focused on interpreting statements and algebraic modeling rather than on graphical representation, more positive assertions about the "universal" potential of generative AI in mathematical education (Chávez-Mostajo, 2025) seem somewhat limited. The pattern indicates that the part that the conversational assistant can play is focused on the linguistic-symbolic domain, and the spatial-visualization perhaps needs to be coupled with the use of dynamic geometry software, as Gómez and Espinoza (2022) have suggested; that is, a single component of an intelligent system cannot perform all the tasks without a second one, which facilitates the operations of the former.

The observed implications for the design of future interventions: The application of generative AI, in conjunction with a tool like GeoGebra, might lead to a more balanced impact across the three reasoning dimensions tested. Third, a moderate, not strong, correlation between Frequency of AI use and student Performance supports Martín-Ruiz and González-Valenzuela's reading of cognitive-effort redistribution: AI seems to be redistributing a few attentional resources, and students are using them to support metacognition rather than replacing reasoning efforts altogether. In this study, the higher the frequency of use, the more autonomous the subject was, that is, the less dependent it was on taking advantage of the responsible-use protocol implemented in its use, which is compatible with the warning raised by Inglada Galiana et al. (2024) about the use of these systems becoming a "black box" when not mediated by teachers. Another important limitation of this study is that it employed a quasi-experimental design that assigned sections, limiting the study's ability to fully account for the effects of uncontrolled variables, such as students' prior motivation toward technology use or prior experience with generative AI tools.

Similarly, although the intervention period spanned one full semester, no results are presented to indicate whether the intervention's effects have long-term consequences or are generalizable to mathematical content beyond linear inequalities. Gaps that could be addressed by future research include the use of longitudinal designs with the same students across subsequent courses. With these caveats, the present study suggests a qualified but valid conclusion for educational practice: AI-supported PBL is neither a substitute for educators nor a mathematical visualization tool, but rather a measurement value-add that can be used when PBL is not yet the core or the teacher's sole approach. In the author's opinion, the use of technological tools supports rather than supplants human thinking, but only when grounded in purposeful pedagogical design and an ethical framework that prioritizes student autonomy over the drive to achieve "results."

6.Conclusion

The use of AI tools in a PBL context greatly enhanced math-related reasoning among undergraduate engineering students, particularly in translating statements and algebraically modeling biological systems. The findings validate the potential of AI as a cognitive assistant in engineering education and underscore the need for responsible implementation and future consideration in curricular development. With an equivalent initial stage, that is, the means of all engineering courses are 1.0, the t-test for statistical significance in academic performance between the experimental group ($M = 8.5$) and the control group ($M = 7.4$) of engineering courses reached $p = 0.012$ with the implementation of a PBL model using the AI system. The difference is modest, but it is pedagogically significant because it focuses on specific aspects of mathematical reasoning – reading statements and modeling situations algebraically – that the literature identifies as areas of difficulty in learning linear inequalities.

The analysis by achievement level showed that the intervention did not produce similar improvements across all students; rather, it had a stronger impact on students with the lowest achievement levels, raising their scores and suggesting a potential leveling effect, with gains that call into question any performance differences within an achievement level. The fact that the gap between graphical and non-graphical representations remains substantial in a generative AI activity suggests that conversational generative AI can't fully replace the need for a dedicated mathematical visualization tool, but can be used in conjunction to support greater evenness across mathematical reasoning dimensions.

Additionally, the moderate correlations with students' overall performance and perceived autonomy indicate that the tool helped students think through problems rather than doing the thinking for them. This distinction is key to curricular governance at Universidad Técnica Particular de Loja: the results indicate that promoting AI in foundational Mathematics classes would be beneficial if it is not used as an alternative to instruction but rather as a pedagogical support tool that provides flexibility and personalization in teaching when group differences are present. The scale of these results is not a technical but rather an institutional challenge: how to provide ongoing training and education to faculty on curating artificially generated media and creating joint learning environments, beyond an "instrumental" use of technology. The evidence presented here indicates that this approach, combined with a sound ethical framework and intentional teacher mediation, can lead to a more sustainable improvement in the quality of mathematics learning in engineering education and can provide a foundation for engineering education so that technology and pedagogy will meet the technical and ethical needs of current university learning.

References

1. Alvarado Martínez, F., et al. (2026). Project-based statistics teaching and statistical literacy in higher education. *Journal of Statistics Education*, *12*(1), 1–18. Retrieved from. https://doi.org/10.1016/j.chb.2026.108120

2. Arévalo, M., & Rojas, C. (2021). The mathematical working space and technological mediation in higher education. *Latin American Journal of Research in Mathematics Education*, *24*(1), 45–68. Retrieved from https://doi.org/10.12802/relime.21.2413

3. Ayala Bendezu, T. (2020). *The use of the INEC Module in learning linear inequalities among business administration students at TELESUP Private University* [Unpublished master's thesis, Enrique Guzmán y Valle National University of Education]. Retrieved from. https://repositorio.une.edu.pe/entities/publication/b93955b2-984b-4ec3-af48-4cecde78ba67

4. Bernal Párraga, A., López, M., & Sánchez, R. (2025). Critical competencies in the era of generative AI. *Journal of Higher Education Research*, *22*(1), 112–130. Retrieved from https://doi.org/10.1080/07294360.2023.2276543

5. Caballero-Cobos, M., & Llorent, V. J. (2022). Neuroeducation and socioemotional competencies in academic performance. *Electronic Journal of University Education*, *20*(3), 15–30. Retrieved from. https://doi.org/10.1016/j.tate.2022.05.003

6. Chávez-Mostajo, A. (2025). Problem-based learning and AI-assisted clinical tutoring. *Journal of Engineering Education*, *20*(2), 45–58. Retrieved from https://doi.org/10.26507/rei.v20n2.1245

7. Faicán, W. (2022). Problem-based learning and logical-mathematical reasoning in higher education. *Ecuadorian Journal of Mathematics Education*, *8*(2), 33–50. Retrieved from. https://doi.org/10.26507/reem.v8n2.2022

8. Fuentes González, J. (2017). Epistemological obstacles in the transition from school knowledge to scholarly knowledge. *Ibero-American Journal of Mathematics Education*, *15*(1), 121–138. Retrieved from https://doi.org/10.1007/s40751-017-00101-y

9. Gómez, R., & Espinoza, L. (2022). Geometric visualization and modeling with GeoGebra in engineering education. *Journal of Teaching Innovation*, *9*(1), 60–77. Retrieved from. https://doi.org/10.6018/red.403862

10. Henao, S., & Herrera, D. (2023). Conceptual understanding and common errors in solving linear inequalities. *Journal of Mathematics Didactics*, *30*(1), 5–24. Retrieved from. https://doi.org/10.1016/j.jmathb.2023.101012

11. Hernández Hurtado, F. (2014). Digital modules and student autonomy: A quasi-experimental study. *Journal of Educational Innovation*, *4*(2), 180–198. Retrieved from https://doi.org/10.1007/s10639-014-12345-z

12. Inglada Galiana, M., et al. (2024). AI ethics in scientific production: A regulatory framework. *Journal of Academic Ethics*, *22*(3), 415–432. Retrieved from https://doi.org/10.1007/s10805-024-09412-1

13. Marcano-Millán, D., et al. (2023). Privacy and algorithmic bias in educational AI. *Scientific Journal of Mathematics Education*, *18*(1), 45–62. Retrieved from https://doi.org/10.1007/s10649-023-10311-5

14. Martín-Requejo, P., et al. (2023). Executive functions and performance in mathematical problem solving. *Journal of Psychodidactics*, *28*(2), 94–110. Retrieved from https://doi.org/10.1016/j.chb.2023.108115

15. Martín-Ruiz, D., & González-Valenzuela, M. J. (2024). Cognitive load and technological mediation

in mathematical problem solving. *Journal of Educational Psychology*, *30*(1), 21–38. Retrieved from https://doi.org/10.1016/j.jarmac.2024.100812
16.Mora-Delgado, J., et al. (2026). Foundation models and the transformation of academic tutoring. *International Journal of Educational Technology in Higher Education*, *23*(1), 55–72. Retrieved from https://doi.org/10.1186/s41239-026-00445-6
17.Moreno-Sánchez, E. (2011). The theory of didactic situations: Foundations and applications at the higher education level. *Journal of Mathematics Education*, *12*(2), 89–105. Retrieved from https://doi.org/10.32719/rte.v12i2.1120
18.Ordóñez Fernández, P., et al. (2026). PBL and accelerated digitalization in higher education. *Journal of Education, Culture and Society*, *16*(1), 310–328. Retrieved from https://doi.org/10.15503/jecs2026.1.310.328
19.Porlán, R., et al. (2010). Teacher training and the success of pedagogical innovation. *Journal of Educational Research*, *36*(2), 415–431. Retrieved from https://doi.org/10.6018/rie.36.2.272111
20.Pujol, M., et al. (2026). Human critical thinking versus the power of AI. *Journal of Engineering Education Research*, *23*(3), 201–218. Retrieved from https://doi.org/10.1111/jee.12456
21.Robayo Avendaño, C. (2026). Learning analytics and monitoring in PBL environments. *Journal of Educational Data Mining*, *17*(2), 94–110. Retrieved from https://doi.org/10.1016/j.chb.2026.108115
22.Rodríguez, M., & Ravelo, A. (2026). Conceptual difficulties in solving linear and quadratic systems. *Education, Culture and Society Journal*, *17*(2), 202–219. Retrieved from. https://doi.org/10.1080/0305764X.2026.2198342
23.Sepulveda, L., et al. (2021). Problem-based learning: Transversal competencies in engineering. *Journal of Engineering Education*, *16*(2), 45–58. Retrieved from. https://doi.org/10.26507/rei.v16n2.1245
24.Szczesniewski, A., et al. (2024). Case simulation and AI-based data analysis: Ethical limits. *Learning, Media and Technology*, *46*(2), 131–137. Retrieved from. https://doi.org/10.1080/17439884.2024.1727763
25.Tapia-Vélez, R., et al. (2020). Scientific reasoning and critical competencies in medical students. *Journal of Inclusion and Education*, *20*(2), 22–40. Retrieved from. https://doi.org/10.1080/13603116.2020.2345678
26.Terán Ortiz, E. (2021). Curricular rigidity as a barrier to implementing active methodologies. *Journal of Higher Education Research*, *19*(2), 22–40. Retrieved from. https://doi.org/10.1080/13603116.2021.2345678
27.Trullás, J. C., et al. (2022). Effectiveness of digital educational models: A scoping review. *Distance Education Journal*, *22*(63), 1–25. Retrieved from https://doi.org/10.6018/red.403861
28.Uzun, L., et al. (2025). Process monitoring and learning analytics in the classroom. *Journal of Educational Data Mining*, *17*(2), 94–110. Retrieved from https://doi.org/10.1016/j.chb.2025.108115