Revising with AI: Undergraduate Students' Perceptions and Trust in AI-generated Feedback on Geometric Proofs

Aldrich Agad[1]
asagad@up.edu.ph

Lady Angela Rocena[2]
lmrocena@up.edu.ph

Joseph Ma. Steven Cabalo[3, 4]
joseph.cabalo@deped.gov.ph
joseph.cabalo@student.ateneo.edu

Christine Nicole Victorio[4, 5]
christine.victorio@student.ateneo.edu
cnvictorio@nu-moa.edu.ph

University of the Philippines Los Baños[1]
University of the Philippines Integrated School[2]
Department of Education Schools Division of Lipa City[3]
Ateneo de Manila University[4]
National University, Philippines[5]

Abstract: *This study investigated undergraduate students' perceptions of ChatGPT-generated feedback on mathematical proofs in projective geometry. Six senior-year BS Mathematics students participated in a problem-solving task, receiving Artificial Intelligence (AI) generated feedback before revising their proofs. Using the lens of the SIPE-AI framework, post-task interviews were analyzed, which revealed that students found the feedback confusing, primarily due to ChatGPT's focus on structure rather than mathematical accuracy. Many perceived the feedback as redundant or misaligned with their professors' expectations, leading to low trust in the feedback generated. While some students acknowledged AI's usefulness in concept recall, most preferred traditional learning resources such as textbooks and instructor feedback. The findings highlight concerns regarding AI's effectiveness in mathematical proof evaluation, emphasizing the need for improvements in AI's ability to provide precise, pedagogically relevant feedback. This study underscores the importance of aligning AI-generated responses with academic expectations in mathematics education.*

Keywords: ChatGPT, AI-generated feedback, student perceptions, mathematical proofs

1. Generative AI in Education

Artificial Intelligence (AI) is transforming contemporary classrooms, particularly in mathematics, by enhancing personalized learning experiences and providing real-time support [1]. By adopting approaches like AI-tutor and AI-simulator, educators can enhance student engagement and understanding [2]. However, this must be approached responsibly, addressing ethical concerns such as biases and equitable access to resources [3], especially in this age where generative pretrained transformers (GPT) such as ChatGPT emerge in the educational landscape. This evolving role of AI in education is supported by studies that indicate a growing comfort among mathematics-oriented students with AI tools like ChatGPT, believing it positively impacted their learning and increased classroom participation and engagement [4, 5, 6].

AI's transformative role in mathematics education has been highlighted in previous research; it offers personalized learning and intelligent tutoring systems, enhancing student engagement and creativity, and automation and personalization. Intelligent tutoring systems use AI to provide real-time feedback and adapt instruction based on students' strengths and weaknesses, enhancing learning efficiency and engagement [7]. Another study noted that AI tools have been shown to enhance positive-activating emotions, situational interest, and self-efficacy in students, thereby reducing cognitive load and fostering a superior learning experience [8]. In addition, for automation, the Virtual AI Teacher system, for instance, autonomously analyzes and corrects student errors, significantly improving learning efficiency and reducing educational costs [9].

While the increasing integration of Large language models (LLMs) like ChatGPT continues to gain interest in educational settings regarding its impact on student learning [10], its use has sparked concerns about over-reliance on these technologies, potentially undermining critical thinking skills[11]. This can result in the acceptance of misleading or incorrect information, as AI systems are prone to generating hallucinated content [12, 13]. Moreover, "Algorithm Aversion" is another key highlight in the use of LLMs, as users tend to exhibit distrust in AI given the tool's accuracy after observing errors in its generated response[14]. Effective AI integration thus involves developing critical source evaluation skills and promoting a balanced approach to enhance students' knowledge and skills development[15], which other studies call for further research especially for Generation Alpha students[16].

2. Students' Interaction with AI-generated Feedback

AI-generated feedback can significantly influence students' learning processes, particularly in areas requiring proof and logical reasoning, such as geometry. [17] proposed an emergent framework, Students' Interactive Proving Experience with AI (SIPE-AI), which outlines how students engage with generative AI when constructing mathematical proofs. This framework, shown in Figure 1, identifies critical processes, such as identifying desired outcomes, revising prompts, and evaluating AI-generated responses, which are essential for students to develop their problem-solving skills. The SIPE-AI framework emphasizes the interactive nature of learning with AI, suggesting that students must take an active role in assessing the validity and relevance of the feedback they receive.

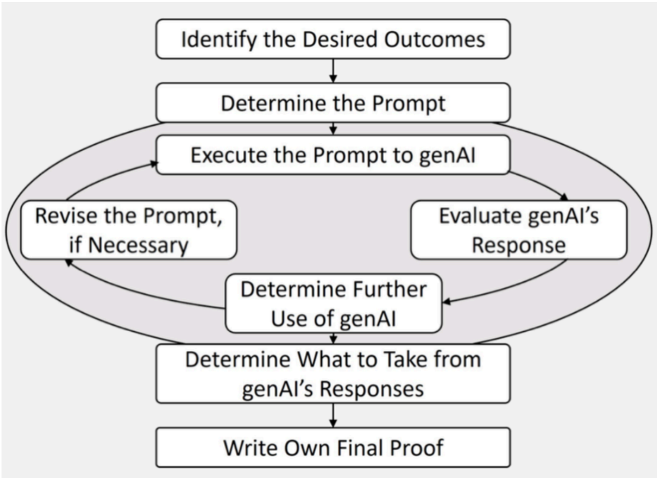

Figure 1 Students' Interactive Proving Experience with AI (SIPE-AI) Framework [17]

Moreover, the framework highlights three key factors shaping students' engagement with AI: conceptions of proof, conceptions of generative AI, and ethical considerations surrounding its use. Students' conceptions of proof influence how they evaluate AI feedback, as they seek responses that resonate with their understanding of rigorous mathematical argumentation. Similarly, their perceptions of AI inform their trust in the feedback provided, impacting their willingness to revise their proofs based on AI suggestions. Ethical considerations also play a role, as students grapple with the implications of using AI in their learning processes, raising questions about academic integrity and the balance between utilizing technological aids and fostering independent mathematical thinking.

We adapt this framework in this paper. However, we will be focusing solely on the perception of the students based on the feedback given to them by ChatGPT, which is generated from a standardized prompt used by all participants. We analyze the responses to identify and compare patterns in students' proof-solving approaches before and after receiving assistance from ChatGPT. We aim to gain insights into the impact of AI-assisted feedback on students' learning behaviors and their perceptions of the problem-solving process.

3. Research Questions

This study, extending from the research of[18], explores students' perceptions of AI in mathematics, focusing on its use in assessments, trust in AI-generated feedback, and its benefits and limitations in geometric proving. Specifically, it aims to answer: (1) What are the undergraduate students' perceptions of AI-generated feedback on mathematical proofs, specifically, about projective geometry? (2) To what extent do undergraduate students trust AI-generated feedback on mathematics assignments and activities? (3) How do undergraduate students revise their proofs in response to AI-generated feedback in projective geometry tasks?

4. Methods

This study explored the undergraduate students' perception of ChatGPT-generated feedback in mathematics through a qualitative research design, adapting a simplified version of the SIPE-AI framework by [17]. The inner circle of the SIPE-AI framework shows the iterative loop students often go through when working with AI during proof construction. In our study, we kept the cycle of prompt generation in the SIPE-AI framework constant. This was a deliberate choice based on the feedback proposed by [19], which we found to be thoughtfully designed and aligned with our study goals. Since our focus was on understanding how students perceive and trust AI-generated feedback, rather than how they generate prompts, we adopted the recommended setup from [19] to maintain consistency and ensure the reliability of our findings. The framework employed in this study is shown in Figure 2.

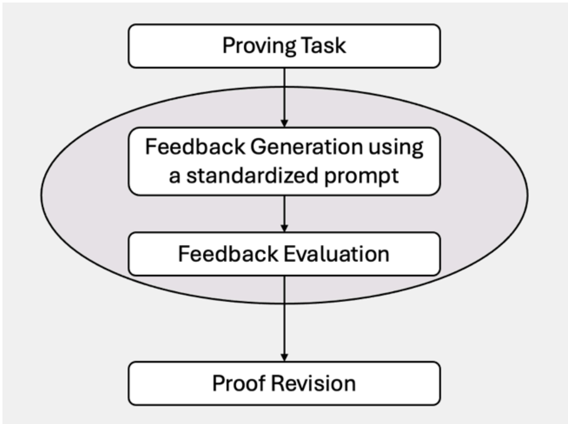

Figure 2 Research framework based on the SIPE-AI Framework

The present study began with a problem-solving task where participants were given the axioms of an affine plane and asked to prove a specific theorem using only these axioms. This task ensures that all participants engage with the problem at the same level, eliminating external influences such as prior exposure to the proof. The topic of the exercise, shown below, focused on Projective Geometry, a core senior-year subject in the BS Mathematics curriculum.

The Proving Task

Prove the statement, **"An affine plane contains at least four points,"** using the following axioms.

Axiom A1: Given two distinct points P and Q, there is one and only one line that contains both P and Q.
Axiom A2: Given a line l and a point P not on l, there is one and only one line m that passes through P and is parallel to l.
Axiom A3: There exist three non-collinear points.

Once participants completed their proof, they submitted their solution to ChatGPT using a standardized, templated prompt that is modified from the template used by [19]. The prompt, shown in Figure 3, includes four key parts: a context that frames the AI as a professor responding to a student in a geometry course; the proving task that asks the student to prove a statement using specific axioms; a space for the student to input their draft proof; and a detailed instruction asking the AI to provide feedback by identifying errors, explaining them, and suggesting improvements without giving the final answer.

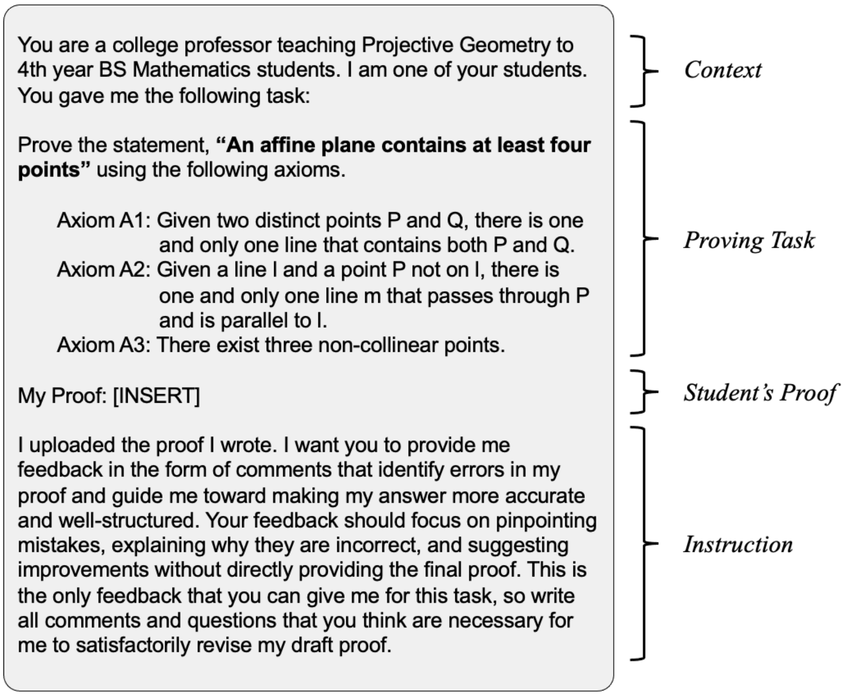

Figure 3 Standardized Prompt adapted from [19]

ChatGPT then generated feedback evaluating the correctness, clarity, and logical structure of the proof, which also provided suggestions for improvement. After receiving the ChatGPT-generated feedback, participants had the opportunity to revise their proof based on the feedback they received. They were given the options to accept, modify, or reject the AI’s suggestions, depending on their assessment of its relevance and accuracy. This step allowed researchers to analyze how students engage with AI feedback and whether it influences their problem-solving and revision strategies.

Following the revision phase, participants completed a qualitative post-task interview that explicitly probed their behavioral engagement with feedback, including questions such as “How did you use the feedback to revise and improve your initial answer?”, “Would you consider using AI-generated feedback in future mathematics tasks? If so, how?”. While these questions aimed to capture participants’ reported actions and intentions during revision, the study relied solely on self-described behaviors rather than direct observation or analysis of their actual revisions.

Data were collected from seven students enrolled in AY 2024-2025 who had completed the prerequisite Euclidean Geometry course. These students are coded Janice, Chad, Karl, Bernadeth, Mela, Gab, and Dwight in this study. Their selection ensured a strong foundation in proof techniques while presenting them with a reasonably challenging problem to assess their proving skills. Participants provided informed consent in compliance with the Philippines’ Republic Act No. 10173, otherwise known as the Data Privacy Act of 2012.

To analyze the data, two approaches were used. First, a thematic analysis of the post-task interview responses was carried out to explore students’ perceptions, trust, and decision-making processes when interacting with ChatGPT's feedback. Second, a comparative analysis was conducted between each student's original and revised proofs to examine the nature of the revisions. This comparison helped identify patterns in how students engaged with and responded to the AI-generated feedback. Together, these analyses provided a comprehensive view of both the observable changes in students' written work and the underlying beliefs and attitudes that informed those changes.

5. Results and Discussion

This study explored undergraduate students' engagement with ChatGPT-generated feedback in projective geometry proof-writing. Thematic analysis revealed that many students found the feedback confusing, as it evaluated their proofs line by line without assessing the overall logical structure, often resulting in contradictory comments. Some students felt the feedback merely reiterated or overexplained their existing arguments, making their proofs more complex than necessary. Others noted that the focus on grammar and sentence construction was unhelpful, given the preference for a specific proof-writing style in mathematics. Despite these issues, a few students found the feedback useful in improving the flow and clarity of their arguments. Overall, most participants expressed low trust in ChatGPT's feedback and preferred more reliable sources such as textbooks, professors, or educational videos. Nevertheless, some recognized its value in reviewing concepts or offering initial guidance for constructing proofs.

These findings, analyzed through the lens of the Students' Interactive Proving Experience with AI (SIPE-AI) framework [17], reveal tensions between AI's capabilities and students' expectations. Three key themes emerged, aligned with SIPE-AI's dimensions: (1) evaluation of AI feedback, (2) conceptions of proof and AI, and (3) decision-making in proof revision.

5.1 Evaluation of AI-generated Feedback

The study shows mixed feelings in students' evaluation of ChatGPT's feedback. While approaching the tool with initial openness, participants consistently encountered structural limitations that undermined its utility. The most significant barrier emerged from ChatGPT's fragmented analytical approach. One student, Janice, clearly articulated this: "It assessed my proof per line, not the flow... some feedback it gave was contradictory." This evaluation is not aligned with students' fundamental conception of mathematical proofs as integrated logical structures where the connections between statements matter more than individual components.

The problem was compounded by ChatGPT's tendency to generate contradictory feedback or merely paraphrase students' existing arguments without substantive improvement. Janice also noted this contradiction, mentioning that she was confused because she didn't know "if the feedback that ChatGPT gave was supporting each other, since some feedback it gave was contradicting." Another student, Karl, echoed this sentiment, noting, "It just reiterated what I mentioned in the proof." This highlights how the AI often failed to provide meaningful advancement of its mathematical reasoning. These limitations significantly eroded trust, with most students concluding they couldn't rely on ChatGPT for rigorous proof validation. Notably, this distrust extended beyond mathematical content to the tool's basic reliability, as several participants reported receiving different feedback for identical proof submissions at different times.

5.2 Conceptions of Proof and AI: Trust and Utility

Students were often frustrated by a mismatch between the feedback they needed and what ChatGPT provided. While they valued logical structure and axiomatic precision, the AI focused instead on grammar and formatting. Gab noted the feedback was “not more so on the correctness of the proof. . . but more so on the grammar,” while Dwight complained it emphasized “structure and notations” rather than validity. For many, this made ChatGPT’s feedback feel irrelevant in a context like projective geometry, where precise application of axioms is crucial.

Despite these limitations, some students found strategic ways to use the tool. Chad used it to improve transitions and conclusions, Karl valued reminders to state assumptions like distinctness and non-collinearity, and Dwight acknowledged its help in recalling forgotten details. Such selective use reflects an adaptive strategy: while distrusting ChatGPT for proof validation, students used it for phrasing, clarity, and memory support, treating it as supplementary rather than authoritative.

Underlying this strategy was a lack of trust in AI’s mathematical accuracy. As Mela put it, they considered feedback only “with a grain of salt since not all of its answers are certain.” Dwight similarly warned, “it is prone to committing errors.” This aligns with “algorithm aversion” [14], where observed mistakes foster skepticism. Students thus engaged critically with AI, integrating its input only when it aligned with their own reasoning, much like using a spell-checker in technical writing: helpful for style, but not for validating correctness.

5.3 Decision-making in Proof Revision

The revisions of the student proofs can be classified into two themes: changes in how the proofs were communicated and changes in the logical reasoning itself. While both themes appeared in the students’ work, the more prominent revision was in how the students expressed and structured their arguments. This suggests that while AI-generated feedback had a limited influence on students’ core mathematical reasoning, it proved more supportive in improving the mathematical communication of their proofs. The observable changes in their revised proofs, however, were heavily influenced by students’ perceptions of ChatGPT’s credibility and usefulness, leading to varied approaches to integrating feedback.

Communicative and Formalization Improvements. Many students, despite expressing frustration with AI’s focus, found its feedback beneficial for refining the articulation of their proofs. They already had the basic logical framework in place, but needed to improve how they presented that reasoning clearly and coherently. For example, as earlier illustrated, Chad found ChatGPT’s feedback most helpful for restructuring his proof and improving its overall coherence, particularly by clarifying transitions between steps and adding conclusions to make his arguments more logical and less confusing. He noted that the AI helped him connect the different parts of his proof, thus enhancing its clarity and formal wording. These revisions are highlighted in bold face in Table 1.

TABLE 1 Chad's Original and Revised Proof

Chad's Original Proof	Chad's Revised Proof
For the axiom A1, assume that there are distinct points, say P and Q, that contain a line, say k. For the axiom A2, assume that there is another point, say R, and a line, say l, not on the point R, there is only one line, say m, parallel to the line l passing through the point R and parallel to the line l. For the axiom A3, we need to show that there are three non-collinear points, which is like the definition of a triangle. We assume that the points P, Q, R are the three non-collinear points for this. With this, we assume that there exists a line that passes through point P and R to make the three points non-collinear. With this, it consists 3 points and 3 lines, but has 4 cases. Another case is if the distinct points are P and a point S that consists of another line say n. And it is parallel to a line that consists of points Q and R. For the axiom A3, there will be additional sets of the three non-collinear points, which are: {(Q,R,S), (P,R,S), (P,Q,S)}. With the sets of non-collinear points, the possible points are P, Q, R, and S. While the possible lines are PQ, PR, PS, QR, QS, and RS. It proves that it is possible to have 4 points and 6 lines in an affine plane.	**Let P and Q be the distinct points, with the axiom A1, there exists exactly one line, say k, that contains both P and Q.** **Moving to the axiom A2,** assume that there is another point, say R, and a line, say l, not on the point R, there is only one line, say m, parallel to the line l passing through the point R and parallel to the line l. For the axiom A3, we need to show that there are three non-collinear points. We assume that the points P, Q, R are the three non-collinear points for this. With this, we assume that there exists a line that passes through points P and R to make the three points non-collinear. With this, it consists 3 points and 3 lines, but there are more cases. Other cases are if the distinct points are P and a points S that consists another line say n **by drawing a parallel line** to one of the lines through P, Q and R. For the axiom A3, there will be additional sets of the three non-collinear points, which are: {(Q,R,S), (P,R,S), (P,Q,S)}. With the sets of non-collinear points, the possible points are P, Q, R, and S. While the possible lines are, PQ, PR, PS, QR, QS, and RS. It proves that it is possible to have 4 points and 6 lines in an affine plane.

Similarly, Karl noted that while ChatGPT "just reiterated what I mentioned in the proof," it "did raise some points that I haven't considered. For example, stating that the point should be distinct and non-collinear. I haven't stated that." As a result, his revision included this formalization. His original proof began, "Suppose P, Q, and R are the three distinct points." In his revised proof, he explicitly added, "Suppose P, Q, and R are the three distinct **noncollinear** points." This addition addresses a foundational assumption in affine geometry and ensures the proof is built on explicitly stated conditions. There is no other change in his proof apart from this.

Bernadeth also focused on elaboration, noting, "the feedback provided by ChatGPT on my proof was almost more about elaboration of the terms or the arguments that I used there. And it's more on the technical terms that we can improve." Her revised proof, presented in Table 2, shows a significant expansion of her arguments for clarity.

TABLE 2 Bernadeth's Original and Revised Proof

Bernadeth's Original Proof	Bernadeth's Revised Proof
Let A be an affine plane and consider 3 non-collinear points in A; say X, Y, Z. Now, by A1, two distinct points are contained in exactly one line; XY, XZ, YZ. Consider the line XY and a point Z not on XY. By A2, there must be exactly one line that passes through Z and is parallel	By Axiom A3, there exist three non-collinear points in the affine plane. Let these points be X,Y, and Z, and by the non-collinearity condition, no single line contains all three points. Now, consider the lines determined by these points. By Axiom A1, we know the following lines exist: Line XY

to XY. Then, consider the line XZ and a point Y not on XZ. By A2, there must be exactly one line that passes through Y and is parallel to XZ. In this manner, the line that passes through Z and is parallel to XY and the line that passes through Y and is parallel to XZ intersect. And in plane, the intersection of two lines is a point, say W. Here, W is distinct and must be non-collinear to X,Y,Z (because of a contradiction of A3). Hence, there exists 4 points in a plane; X, Y, Z, W.	passing through points X and Y, Line XZ passing through points X and Z, Line YZ passing through points Y and Z. Next, consider the line XY, and note that Z is not on this line (since the points are non-collinear). According to Axiom A2, there must exist a unique line through Z that is parallel to line XY. Let's call this line l2. Similarly, consider the line XZ, and note that Y is not on this line. By Axiom A2, there must exist a unique line through Y that is parallel to line XZ. Let's call this line l2. Now, because the lines l1 and l2 are parallel to lines XY and XZ, respectively, and both are in the same affine plane, they must intersect at a single point. Let's denote this point of intersection as W. Since W lies on both l1 and l2, and is distinct from X,Y, and Z (because W is not collinear with any two of X, Y, Z), we conclude that W is a fourth distinct point in the affine plane. Thus, we have shown that an affine plane must contain at least four distinct points: X, Y, Z, and W.

Logical and Structural Refinements. While less common, other students made important adjustments that improved the logical rigor or structural coherence of their proofs. Despite Janice's experienced confusion from ChatGPT's line-by-line feedback, she revised her proof by "remov[ing] the parts that ChatGPT told me were wrong" and following her "gut feeling." Her original proof included an ambiguous "Case 1" and "Case 2" analysis that complicated her argument for finding a fourth point. Her revised proof, as shown in Table 3, eliminated this problematic case analysis by restructuring the argument around triangle formation and a more direct application of Axiom A2 to create parallel lines and ensure an intersection. This change significantly improved the structural soundness and logical flow of her proof by replacing an unclear conditional argument with a more direct constructive approach based on the axioms.

TABLE 3 Janice's Original and Revised Proof

Janice's Original Proof	Janice's Revised Proof
Applying Axiom A3, assume there exist three non-collinear points, say P, Q, and R. Consequently, these points are distinct from each other. By Axiom A1, we'll form line PQ, QR, and PR. Assuming there is a line l and it does not contain any of the points P, Q, and R (by axiom A2). We'll consider two cases. Case 1: There is no line passing through the given points that is parallel to l. This means that there is a point of intersection between the lines PQ, QR, and PR. Consequently, there will be at least 4 points. Case 2. I have to show here that it'll never be the case that all lines PQ, QR, and PR are parallel to l, since it	Applying Axiom A3, assume there exist three non-collinear points, say P, Q, and R. By Axiom A1, we'll form three distinct lines PQ, QR, and PR, which are non-parallel, forming a triangle. Consider a line, say l, that does not contain any of the points P, Q, and R. By Axiom A2, there exists a line parallel to l passing through a point not on l. Without loss of generality, let's say line PQ is parallel to line l. Since P, Q, and R are distinct non-collinear points, it'll be the case that lines QR and PR are not parallel to line l, forming a point of intersection. Which proves there is at least 4 points in an affine plane.

will mean that points P, Q, and R are not collinear. So, at some point, there will be an intersection, forming the 4th point.	

Mela's revised proof, presented in Table 4, also shows a notable improvement in the presentation of a contradiction, despite her expressed algorithm aversion. Her original proof's contradiction was stated somewhat abruptly. In the revised version, this is presented with much greater clarity and logical precision. This refinement enhanced the logical rigor and conciseness of her argument by making the flow of the contradiction more explicit and removing extraneous details.

TABLE 4 Mela’s Original and Revised Proof

<table>
<tr><th>Mela’s Original Proof</th><th>Mela’s Revised Proof</th></tr>
<tr><td>From Axiom 3, there exists three non-collinear points P,Q,R. By Axiom 1 there is a unique line QR. From Axiom 2 there is a line l containing P parallel to QR. There is also a unique line PQ. Since R is an element of PQ, there is a unique line m containing R and parallel to PQ. Note that l is not parallel to m. By contradiction PQ || m || l || QR so PQ || QR. This is a contradiction since PQ != QR and Q = PQ intersection QR != empty. This means that there is a point S = m intersection l since s is an element of m and m || PQ and m != PQ then S is not an element of PQ. Hence S !=P and S!=Q Also, S != R and S is a fourth point. Therefore, there are four points P, Q, R, and S</td><td>By Axiom A3, there exist three non-collinear points P, Q, and R. By Axiom A1, there is a unique line QR passing through Q and R. By Axiom A2, since P is not on line QR, there is a unique line l passing through P that is parallel to QR. Next, consider line PQ. Since R is on PQ, there exists a unique line m through R parallel to PQ by Axiom A2. If lines l and m were parallel, we would have PQ || m ||l || QR, implying that PQ || QR. But this contradicts the fact PQ != QR. Therefore, lines l and m are not parallel and must intersect at point S. Since S is distinct from P, Q, and R, S is a fourth point in the affine plane.</td></tr>
</table>

Rejection of Feedback and Perceived Utility. Not all students chose to integrate ChatGPT's feedback into their revisions. Dwight, for instance, stated, "Well, technically I did not improve my proof because as per ChatGPT it was already correct, it just needs more elaboration on some things as but personally I don't want that because it would just make my proof more complicated. I want it to be as straightforward as possible." This highlights a tension between AI's tendency to provide elaboration and some students' preference for concise, direct proofs, particularly when they already felt their core logic was sound.

Similarly, Gab noted that "most of the corrections provided by ChatGPT focus on my use of notations and, like the clarification of my sentences, I did not really use the feedback given by ChatGPT." While he revised his proof, the changes are subtle and focused on minor rephrasing for clarity rather than substantial alterations.

These varying approaches to revision demonstrate the critical role of student perception and trust in AI-generated feedback. Students were more willing to engage with feedback that aligned with their understanding of what constitutes a "good" proof, whether that was a more formal structure, clearer transitions, or a more rigorously stated contradiction, while often rejecting suggestions they felt were unnecessary elaborations or stylistically misaligned with their goals. This selective engagement underscores that AI tools may currently serve best as supplementary resources for proof refinement rather than primary validation mechanisms, with their effectiveness heavily dependent on students' ability to critically evaluate and strategically employ the feedback within their existing mathematical understanding.

6. Conclusion and Recommendations

This study explored undergraduate students' perceptions of AI-generated feedback in projective geometry proof writing. While ChatGPT provided structured responses, most students found them confusing and of limited utility, preferring textbooks or instructor guidance. Confusion often arose from ChatGPT's line-by-line evaluations, which overlooked overall logical flow and sometimes offered contradictory suggestions. A minority of students noted supplementary benefits, such as reinforcing concepts, improving transitions, and clarifying assumptions. For these students, AI feedback enhanced communicative clarity but rarely strengthened core reasoning. Overall, students remained skeptical, criticizing fragmented analysis, contradictions, and deviations from disciplinary norms emphasizing concision and logical rigor. Students' engagement reflected both potential and limits of AI tools: they selectively applied suggestions, treating the feedback as supplementary rather than authoritative. Revisions were mostly communicative, improving clarity, coherence, and alignment with formal style, refining transitions, explicitly stating assumptions, and clarifying terms, while substantive logical corrections were less common. Trust in AI strongly influenced adoption: low credibility reduced willingness to act on style-focused feedback. This underscores that revised proof reflects not just the AI's suggestions but also students' evaluative judgments.

To enhance educational value, AI tools should be scaffolded by instructors who guide critical evaluation and application of feedback. Paired with human instruction, AI can support both revision and reasoning. Future research may explore interactive or adaptive systems capable of holistic, dialogic proof evaluation, addressing the mismatch between students' need for integrated guidance and current systems' piecemeal feedback.

References

[1] A. Adair, Teaching and learning with AI: How artificial intelligence is transforming the future of education, *XRDS Crossroads ACM Mag. Students* **29** (2023), no. 3, 7–9. Available at https://dl.acm.org/doi/pdf/10.1145/3589252.

[2] A. Agad, J. M. S. Cabalo, L. A. Rocena, and C. N. Victorio, Perceived impact of AI-generated feedback on undergraduate mathematics learning in geometric proof, *E-Proceedings 9th ICMI-East Asia Regional Conf. Math. Educ.*, Seoul Nat. Univ., Korea, 2025.

[3] M. Alashwal, Empowering education through AI: Potential benefits and future implications for instructional pedagogy, *Pupil: Int. J. Teaching, Educ. Learning* (2024), 201–212. Global Research & Development Services.

[4] B. J. Dietvorst, J. P. Simmons, and C. Massey, Algorithm aversion: People erroneously avoid algorithms after seeing them err, *J. Exp. Psychol. Gen.* **144** (2015), no. 1, 114–126. Available at https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=2466040.

[5] E. Elstad, AI in education: Rationale, principles, and instructional implications, *arXiv preprint* arXiv:2412.12116, 2024. Available at https://arxiv.org/pdf/2412.12116.

[6] J. Lademann, J. Henze, and S. Becker-Genschow, Augmenting learning environments using AI custom chatbots: Effects on learning performance, cognitive load, and affective variables, *Phys. Rev. Phys. Educ. Res.* **21** (2025), no. 1, 010147. Available at https://link.aps.org/pdf/10.1103/PhysRevPhysEducRes.21.010147.

[7] S. Liang, W. Zhang, T. Zhong, and T. Liu, Mathematics and machine creativity: A survey on bridging mathematics with AI, *arXiv preprint* arXiv:2412.16543, 2024. Available at https://arxiv.org/pdf/2412.16543.

[8] L. Maulida, P. Nurossobah, B. A. Aura, E. D. Nengsih, and R. Rasilah, Improving the effectiveness of mathematics learning through artificial intelligence: Literature review, *J. Gen. Educ. Humanities* **3** (2024), no. 4, 323–338. Available at https://journal-gehu.com/index.php/gehu/article/download/267/282.

[9] P. J. M. Melchor, L. S. Lomibao, and J. O. Parcutilo, Exploring the potential of AI integration in mathematics education for Generation Alpha—Approaches, challenges, and readiness of Philippine tertiary classrooms: A literature review, *J. Innovations Teaching Learning* **3** (2023), no. 1, 39–44.

[10] E. Mollick and L. Mollick, Assigning AI: Seven approaches for students, with prompts, *arXiv preprint* arXiv:2306.10052, 2023. Available at https://arxiv.org/pdf/2306.10052.

[11] T. Nazaretsky, P. Mejia-Domenzain, V. Swamy, J. Frej, and T. Käser, AI or human? Evaluating student feedback perceptions in higher education, *European Conf. Technology Enhanced Learning* (2024), 284–298, Springer Nature Switzerland, Cham. Available at https://osf.io/6zm83/download.

[12] J. L. Patero, Revolutionizing math education: Harnessing ChatGPT for student success, *Math. Educ.* **3** (2023), no. 1, 807–813. Available at https://www.researchgate.net/.../Revolutionizing-Math-Education-Harnessing-ChatGPT-for-student-success.pdf.

[13] J. P. Remoto, ChatGPT and other AIs: Personal relief and limitations among mathematics-oriented learners, *Environ. Soc. Psychol.* **9** (2024), no. 1, 1–13. Available at https://pdfs.semanticscholar.org/0dfe/638b1378b9a597776421dcb545f1da6cb1c2.pdf.

[14] D. Serhan and N. Welcome, Integrating ChatGPT in the calculus classroom: Student perceptions, *Int. J. Technol. Educ. Sci.* **8** (2024), no. 2, 325–335. Available at https://files.eric.ed.gov/fulltext/EJ1426988.pdf.

[15] T. Xu, Y. Zhang, Z. Chu, S. Wang, and Q. Wen, AI-driven virtual teacher for enhanced educational efficiency: Leveraging large pretrain models for autonomous error analysis and correction, *Proc. AAAI Conf. Artificial Intelligence* **39** (2025), no. 28, 28801–28809.

[16] L. Yan, L. Sha, L. Zhao, Y. Li, R. Martinez-Maldonado, G. Chen, et al., Practical and ethical challenges of large language models in education: A systematic scoping review, *Brit. J. Educ. Technol.* **55** (2024), no. 1, 90–112. Available at https://bera-journals.onlinelibrary.wiley.com/doi/pdf/10.1111/bjet.13370.

[17] K. Yatani, Z. Sramek, and C. L. Yang, AI as extraherics: Fostering higher-order thinking skills in human-AI interaction, *arXiv preprint* arXiv:2409.09218, 2024. Available at https://arxiv.org/pdf/2409.09218.

[18] C. Zhai, S. Wibowo, and L. D. Li, The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review, *Smart Learn. Environ.* **11** (2024), no. 1, 28. Available at https://link.springer.com/content/pdf/10.1186/s40561-024-00316-7.pdf.