Social Informatics Journal
Business, Politics, Law, Environmental Sciences & Technology

Vol 12, No 2, 2026 Series
Publication of the Faculty of Computational Sciences, West Midlands Open University, Lagos, Nigeria
Special Issue in collaboration with the International Institute for Multidisciplinary and Development Research

The Impact Of AI-Based Tutoring Systems On Students' Conceptual Understanding And Proof Skills In Secondary Geometry

[1*]Amao, F.A. (PhD) & [2]Oladapo, D.I. (PhD),
[1, 2]Department of Mathematical Sciences
Adeleke University, Ede, Osun State Nigeria.
E-mails: [1]amao.folake@adelekeuniversity.edu.ng; [2]oladapo.ifeoluwa@adelekeuniversity.edu.ng

ABSTRACT

This quasi-experimental study examined differences in secondary students' geometry conceptual understanding and proof skills associated with the use of an AI-based tutoring system. Sixty-four Year 11 students (two intact classes, randomly assigned to condition) participated. The experimental group ($n = 32$) learned with GeoAI Tutor, an adaptive system providing stepwise feedback, hints, and dynamic visualisations; the control group ($n = 32$) received conventional teacher-led instruction over six weeks. Two researcher-developed instruments measured conceptual understanding (GCUT, $\alpha = 0.86–0.89$) and proof skills (GPST, $\alpha = 0.82–0.85$). Independent-samples t-tests showed the experimental group outperformed the control group on both measures (GCUT: $t(62) = 9.87$, $p < .001$, Cohen's $d = 2.33$ [1.67, 2.99]; GPST: $t(62) = 8.44$, $p < .001$, $d = 1.96$ [1.33, 2.59]). These large effect sizes exceed typical ITS meta-analytic averages ($d \approx 0.50$; Steenbergen-Hu & Cooper, 2014) and should be interpreted cautiously pending independent replication with standardised assessments. The findings suggest that AI-based tutoring systems, when designed to scaffold rather than bypass reasoning, may enhance geometric proof construction. Replication with longitudinal designs and standardised tests is needed.

Keywords: Artificial Intelligence, Intelligent Tutoring Systems, Geometry Teaching, Standards, Conceptual Knowledge, Mathematical Proof, Secondary Mathematics.

Journal Reference Format:
Amao, F.A. (PhD) & Oladapo, D.I. (PhD), (2026) The Impact Of AI-Based Tutoring Systems On Students' Conceptual Understanding And Proof Skills In Secondary Geometry. Social Informatics, Business, Politics, Law, Environmental Sciences & Technology Journal. Vol. 12, No. 1. Pp 1-. A Publication of the Faculty of Computational Sciences, West Midlands Open University, Lagos, Nigeria Special issue in collaboration with the International Institute for Multidisciplinary and Development Research Accessible at www.isteams/socialinformaticsjournal. dx.doi.org/10.22624/AIMS/SIJ/V12N2P1

1. INTRODUCTION

Geometry provides a domain for developing spatial reasoning, logical thinking, and formal proof skills, it's competencies foundational within school mathematics and beyond (Yi et al., 2024). Despite its importance, geometry is widely perceived as difficult by secondary students. Though students can develop procedural mastery of formulae and formulaic problems, many fail to justify why a particular answer to a problem is correct, an issue that represents a larger conflict between

social Informatics Journal
Business, Politics, Law, Environmental Sciences & Technology
Vol 12, No 2, 2026 Series
Publication of the Faculty of Computational Sciences, West Midlands Open University, Lagos, Nigeria
Special Issue in collaboration with the International Institute for Multidisciplinary and Development Research

procedural and conceptual mastery in mathematics instruction. Research indicates that geometry is viewed by most students as a set of unrelated rules instead of a system of thought that is logically organized (Yi et al., 2024). Conventional classroom instruction is often based on a transmission model with the teacher showing students how things are done and the students imitating them. Though this practice can be helpful in encouraging results on exercises of routine, it frequently does not encourage the type of structural analysis and free argumentation that is the basis of real understanding of mathematics. As a result, a lot of students learn proofs without learning to understand the deductive reasoning behind them, thus restricting conceptual growth and undermining deductive reasoning skills.

This is a crisis especially in the area of geometric proof. Proof is key to mathematics since it is the logical explanation of mathematical claims. To build a geometric proof, students would have to recognise connections between the geometric objects and choose the right theorems and structure the ideas in a logical order. It is because many students are unable to make the transition between identifying a pattern, and giving a deductive reason as to why that pattern exists (Wertheimer, 1990). One of the factors is the way in which proof is normally taught: in many cases, teachers often present ready-made proofs to students, which gives them little time to get involved in the process that produces the proofs (Harel & Sowder, 2007; Stylianides and Stylianides, 2009).

An increasing number of studies show that AI could promote learning in mathematics especially when it comes to reasoning tasks. Instant feedback will allow students to identify and rectify errors in the process of solving problems, rather than only after completing the tasks, which helps them modify their strategies and revise their concepts. Gains in conceptual knowledge, engagement, and self-efficacy have been reported (Canonigo, 2024). In geometry, AI-based systems also provide affordances in the form of dynamically visualised tools that enable students to interactively manipulate geometric objects, test conjectures and observe invariant and variant properties, activities that aid learning spatial reasoning and constructing concepts, both of which are precursors to learning proof. Proof construction, which involves the construction of proofs step-by-step, can also be scaffolded by AI-based tutoring systems.

These systems allow students to arrange their arguments, and gain metacognitive control over the process of making a proof, by breaking down proofs into smaller, individual steps, necessitating justification of each step, and directing students through one step at a time. Deductive reasoning could also be enhanced by providing real-time feedback and dynamically adaptive hints (Marwiang et al., 2025). Recent empirical research brings more evidence of the promise of AI in mathematics education. Experimental studies have discovered that AI tutoring systems are able to generate higher learning outcomes than some active learning methods (Kestin et al., 2025). According to other studies, adaptive tutoring has the potential of empowering mathematical reasoning and problem-solving (Menlah & Boateng, 2025). According to meta-analytic reviews, AI interventions have the potential to alleviate mathematics performance in both K–12 environments (Yi et al., 2024). Concurrently, instructional design is critical as emphasized in the literature. AI does not enhance learning automatically; students may simply use system-generated responses in case the learning

social Informatics Journal
Business, Politics, Law, Environmental Sciences & Technology
Vol 12, No 2, 2026 Series
Publication of the Faculty of Computational Sciences, West Midlands Open University, Lagos, Nigeria
Special Issue in collaboration with the International Institute for Multidisciplinary and Development Research

context does not demand independent thinking, which can reduce learning benefits (Bastani et al., 2025). As such, the incorporation of AI in teaching mathematics must be well planned so as to facilitate and not to replace student thinking.

Although there has been an increasing interest in the same domain, it has still been observed that there is a limited number of studies which are specifically based on the geometric reasoning and proof in secondary school education. A good part of the current literature evaluates general mathematics performance or procedural skill, with a relative lack of studies that focus on formal deductive reasoning. In particular, to the best of our knowledge, no controlled studies have yet concurrently investigated the impact on both conceptual knowledge and constructing proofs- two related but different outcomes- in the secondary school geometry class. Moreover, evidence from developing educational contexts remains very limited. Studies in Nigerian secondary schools have indicated encouraging findings, in which AI-enhanced geometry teaching was found to lead to better learning outcomes and student engagement (Akintade and Olaore, 2025), but there are few controlled studies in this context.

Since the issue of the importance of geometric reasoning and failure to learn proofs by many students is still a problem, it is justified to conduct controlled studies to understand whether AI tutoring systems can facilitate a higher level of conceptual understanding, as well as in the development of proof skills than conventional instruction. This study examined the effect of an AI-based tutoring system on secondary school students' conceptual understanding and proof skills in geometry, comparing AI-supported instruction with conventional classroom teaching.

The study addressed the following research questions:

1. Is there a significant difference in geometry conceptual understanding between students who receive instruction through an AI-based tutoring system and those who receive conventional teacher-led instruction?
2. Is there a significant difference in geometry proof skills between students who receive instruction through an AI-based tutoring system and those who receive conventional teacher-led instruction?

H_1: Students in the AI-based tutoring condition will score significantly higher than controls on both conceptual understanding and proof skills post-test.

The aim was to provide empirical evidence on whether intelligent tutoring systems can enhance geometric reasoning and proof construction in secondary mathematics education.

2. LITERATURE REVIEW

2.1 Challenges in Learning Geometry and Proof

The learning of geometry involves students combining visual-spatial thinking and deductive reasoning, which is quite challenging to many learners. The theory of geometrical figure apprehension by Duval (1998) points out four forms of cognitive processing in geometry, namely:

perceptual, sequential, discursive and operative. AI-based scaffolding could theoretically reduce extraneous cognitive load (Sweller, 1988) by structuring stepwise reasoning, though direct empirical evidence in geometry remains limited. In contrast to the other branches of mathematics where the students can rely on the use of calculation as the main tool, geometry requires the student to interpret space relations, define the properties of figures and create reasoned arguments to cause the assertions. Even though the majority of students are able to identify geometric shapes and use formulas about area or angles, procedural knowledge does not always equate to the ability to construct a proof, it involves more than a rote-level procedure (Yi et al., 2024).

Building a proof means that students have to organize the statements in the logical sense, support each of the relations, and organize several properties in one consistent argument. It is a process that requires a conceptual grasp and cannot be done through mere recollection. Studies have indicated that part of the problem faced, as regards proof has been the result of the existing pedagogical practices. In most classrooms, teachers will give theorems and proofs (already done) and students will copy them and memorise. This practice can build up recall, but it is not always able to initiate thinking. The students can learn to think of proof in terms of a template instead of as a description built out of geometric connections, thus undermining the development of deep understanding and constraining real mathematical thought (Harel and Sowder, 2007; Wertheimer, 1990).

The recent research in the field of mathematics education suggests that the acquisition of proofs is enhanced when the students get a chance to investigate the relations of geometry, develop their hypotheses and get feedback on how they reason when solving the problem. There is a significant role of active involvement in geometric structure in developing the deductive reasoning (Yi et al., 2024). It is also applicable to affective factors: students might feel stressed when they have to engage in proof activities, and might lose interest in reasoning when they experience failure. Research into mathematics learning instruction has indicated that systematic aids can overcome such obstacles, and that reasoning tasks can be more accessible with support mechanisms that deliver gradual assistance (Menlah & Boateng, 2025).

2.2 Intelligent Tutoring Systems in Mathematics Education

One way to overcome the problems of proof learning is an intelligent tutoring system (ITS). These computer systems are based on artificial intelligence, and offer some adaptive and individualised guidance, seeking to simulate the assistance of a human tutor. An ITS examines the response of the students in the solving of a problem, detects errors or misconceptions and provides feedback to the learner, based on their needs. Contemporary systems typically incorporate learner models that track student knowledge and diagnose areas of difficulty. The system then changes the choice of tasks or level of support provided accordingly allowing a certain degree of personalisation which would be difficult to accomplish when the same teacher worked with a large number of students at the same time in a classroom. Studies indicate that these types of systems are compatible with mathematics instruction since mathematical reasoning is sequential and cumulative in nature. In cases where a student commits a mistake during an initial step, later inferences can be invalid; instant feedback will enable the learner to identify the mistake and rectify it before it leads to further errors (Lin et al., 2023).

Attempts to aid in proof learning with the use of technology are not new. One of the first computer-based systems to provide students with a guided experience in the construction of proofs was the Geometry Proof Tutor (Wertheimer, 1990) which demonstrated that deductive reasoning with technology-mediated scaffolding was possible. Although early systems demonstrated the feasibility of technology-mediated scaffolding, contemporary AI-based tutoring systems leverage machine learning to offer adaptive, personalised support at scale. Such systems are able to detect errors, provide specific hints, and even lead problem-solving in a more personalised way than most classroom environments can (Lin et al., 2023).

The AI in education reviews show that intelligent tutoring systems are increasingly used in the teaching of mathematics. According to Lin et al. (2023), AI-based learning systems have the ability to predict the behaviour of learners, generate individual learning trajectories, change task difficulty, choose the right exercises, and provide hints to lead students to the right solution. Adaptive system of problem selection, step-by-step feedback, automated hint system and interactive visual tools are common features of mathematics tutoring systems. Positive results are associated with these features and reported in empirical studies. Marwiang et al. (2025) discovered that the systems with real-time feedback enhanced conceptual learning and problem solving in students. Kestin et al. (2025) found out that AI-based tutoring environment could be as effective as traditional teaching when properly designed.

Nevertheless, there are limitations to the ITS literature. Most trials use short-term interventions (usually two to four weeks), which are based on researcher-made tests that might be biased towards the treatment condition, and seldom have long-term follow-ups (Steenbergen-Hu & Cooper, 2014). It is yet unknown whether gains made with AI support will be sustained once the support stops. These restrictions warn against excessive generalisation of findings of any one study and reiterate that there is need to repeat findings using standardised tests and longitudinal follow-up method.

2.3 AI Based Tutoring Systems for Geometry Learning

Geometry is a particularly promising domain for AI-supported learning. Geometric concepts frequently depend on spatial relationships that students can visualise and manipulate. Interactive software allows students to move points, lines, and shapes; observe geometric transformations; and examine how changes to one element affect other parts of a figure. AI-based systems extend these capabilities by combining dynamic visualisation with adaptive instructional support. Not only do students manipulate figures, but also get prompts and feedback associated with their observations. An example of this is that a student may drag the corners of two triangles to verify conditions of congruence.

The system can then suggest to the student what properties are still equal, to state the corresponding congruence theorem, and to prove the conclusion. This combination of observation and reasoning may fortify the relationship between visual explanation and formal demonstration.

Social Informatics Journal
Business, Politics, Law, Environmental Sciences & Technology
Vol 12, No 2, 2026 Series
Publication of the Faculty of Computational Sciences, West Midlands Open University, Lagos, Nigeria
Special Issue in collaboration with the International Institute for Multidisciplinary and Development Research

There is some initial evidence that such interactions can be effective, but there is little domain-specific research in geometry. Canonigo (2024) discovered that AI-assisted platforms enhanced conceptual knowledge and mathematics self-efficacy using interactive activities and providing instant feedback. Xing et al. (2025) conducted a design-based study which found that teachable agents powered by AI improved mathematical knowledge by encouraging students to describe and justify their thinking- a task directly applicable to building proofs. Meta-analytic results also attest to the potential of AI in mathematics education; Yi et al. (2024) found significant increases in mathematics performance in K-12 schools, with more significant effects in schools with adaptive feedback, interactive problem-solving, and guided reasoning.

The role of generative AI in aiding mathematical reasoning has been studied recently as well. It was proposed that these tools might contribute to the students learning the structure of proof and the reasoning when they are used judiciously (Fitri et al., 2025). Wardat et al. (2023) claimed that generative AI can aid in mathematics education by generating explanations and providing alternative strategies, as well as exploration of concepts. Nevertheless, the advantages are dependent on the design of instruction. Bastani et al. (2025) discovered that free access to generative AI can affect learning by making students rely on answers provided by a machine instead of reasoning independently, which is a crucial point, since AI must not be used to substitute this process but to support it. The use of AI in geometry instruction, in Nigerian secondary schools has shown that AI-based tools enhance student engagement and performance in mathematics education (Akintade and Olaore, 2025), and AI-based tutoring can be used to support mathematics learning in a variety of education systems.

2.4 Research Gap

The literature reveals growing interest in AI for mathematics education alongside several notable gaps. This study addressed three specific gaps. First, while prior research has measured general mathematics achievement or procedural skill (e.g., Yi et al., 2024), the present study focused specifically on formal deductive reasoning and proof construction, arguably the most demanding aspect of secondary geometry conceptual understanding. There are currently no recent controlled studies using AI-based adaptive systems have simultaneously examined both outcomes. In the same sample of secondary geometry students. Second, majority of the research uses descriptive or observational designs, which, although effective in determining patterns, do not permit strong causal inferences. The current research used pretest equivalence verification and quasi-experimental design to allow more powerful causal inferences than the descriptive or observational research designs allow. Third, despite the existence of evidence in developing educational contexts (Akintade and Olaore, 2025), controlled assessments of AI-based geometry teaching are few in developing educational contexts, where access to instructional material, teacher experience, and classroom factors may vary significantly compared to those in well-resourced settings.

By comparing AI-supported instruction with conventional teaching on two interrelated outcomes, conceptual understanding and proof skills, this study aimed to determine whether intelligent tutoring systems can improve students' geometric reasoning and formal proof construction beyond what conventional instruction achieves.

3. METHODOLOGY

3.1 Research Design

This study employed a quasi-experimental pretest-posttest control group design to examine the effect of an AI-based tutoring system on students' conceptual understanding and proof skills in geometry. This design was appropriate because it permitted comparison of learning outcomes under two instructional conditions within authentic school settings. Intact classes were randomly assigned to conditions (cluster randomisation), a design that strengthens causal inferences relative to non-randomised quasi-experiments (Shadish et al., 2002), though individual random assignment was not feasible. Quasi-experimental designs are widely used in educational technology research, particularly in studies evaluating intelligent tutoring systems (Lin et al., 2023). Two groups were involved. The experimental group was instructed using an AI-based tutoring system, whereas the control group received the conventional teacher-centred instruction. Both groups were given identical assessments before and after the intervention. The pretest established baseline performance in conceptual understanding and proof skills and was used to verify the comparability of the two groups prior to treatment. The posttest measured learning outcomes following the intervention. By comparing score changes across both groups, the study examined whether the AI-based tutoring system was associated with greater learning gains than conventional instruction.

3.2 Participants

The participants were 64 senior secondary school students (Year 11) enrolled in geometry courses at two public secondary schools in Ogun State, Nigeria. Both schools followed the national curriculum, which includes formal geometric reasoning and proof at this level. Students ranged in age from 15 to 17 years, $M = 16.2$, $SD = 0.8$. The sample was made up of 34 female students (53.1%) and 30 male students (46.9%). Participants were drawn from two intact classes, one assigned to the experimental condition and the other under the control condition. Intact classes were randomly assigned to conditions, to reduce selection bias and maximize group comparability. A priori power calculation with the G*Power (Faul et al., 2009) revealed that, with an independent-samples t-test, 0.05, power = 0.80, and a medium-to-large effect (d = 0.70), 52 participants were needed. The 64-sample size was sufficient to give sufficient statistical power to detect effects of the expected magnitude. The two schools were chosen within the same school district and were quite similar in terms of socioeconomic profile, previous year mathematics examination by grade (mean difference less than 3%), and size of the classroom, which minimized the possibility of school-level variation confounding the outcomes.

All students were given pretest tests before the intervention to determine the level of performance; pretest scores determined that there was no significant difference in pre-intervention performance between the two groups. Participation by the students was voluntary and all the students were informed about the purpose of the study before data collection and both students and their parents or guardians signed the written consent. The confidentiality of the participants was ensured and all the responses were utilized in the course of research alone.

3.3 Instructional Intervention

The intervention was conducted over six weeks, with three 45-minute sessions per week (18 sessions in total). Both groups studied the same geometry content: triangle congruence, triangle similarity, and deductive proof. The groups differed only in the method of instruction.

Experimental group.

The experimental group students were instructed with GeoAI Tutor, an AI-based tutoring system designed by the research team to teach geometry and proof construction. The system used a Bayesian Knowledge Tracing (BKT) algorithm (Corbett and Anderson, 1995) to estimate whether a student had mastered each knowledge part (e.g., "identifying corresponding parts of congruent triangles," "applying the SAS congruence criterion"). Initialisation and updating of the BKT parameters including initial probability of mastery, learning rate, guess rate and slip rate was based on the results of a pilot study (N = 30) and updated after each student response. The architecture of the system had a learner model that monitored mastery of geometric concepts, a pedagogical module that chose the right tasks and level of scaffolds, and a dynamic visualisation interface. It had four main features:

1. Adaptive problem sequencing: The level of difficulty of tasks was varied in response to the performance of the students. Students who demonstrated mastery were given more complex problems that required more reasoning and those that were struggling were given additional practice and prompts to help them understand the problems better.
2. Real-time feedback: This system was able to solve problems immediately with step level feedback and this assisted the students to be able to identify and correct errors when they occurred, and not at the end of the task. The purpose of feedback was to be explanatory rather than simply corrective to reveal the geometry behind it.
3. Hints on the proof tasks: The system did not give complete solutions but instead offered systematic hints on a four-level hierarchy: (a) restatement of problem, (b) identification of applicable, relevant geometric relationships, (c) suggestion of applicable theorems, and (d) indication of next logical step. This scaffold assisted students in knowing what to do next, that is helpful without undermining the necessity to know how to think individually.
4. Interactive geometric visualisations: A drawing tool that is interactive enabled students to adjust geometrical figures, and observe how lengths of side, angles and spatial relationships change in real time. This feature enabled students to make hypotheses and explore how changes affected the properties of geometry, thereby supporting the connection between visual knowledge and formal reasoning.

Instructional activities progressed from concept exploration (weeks 1–2) to guided proof construction (weeks 3–4) to independent proof problems (weeks 5–6). Students worked individually on the system, with the teacher circulating to provide technical assistance as needed.

Control group.

social Informatics Journal
Business, Politics, Law, Environmental Sciences & Technology
Vol 12, No 2, 2026 Series
Publication of the Faculty of Computational Sciences, West Midlands Open University, Lagos, Nigeria
Special Issue in collaboration with the International Institute for Multidisciplinary and Development Research

The control group students were provided with teacher instructions on the same subjects through a traditional classroom setting. Educators described the concepts with direct instructions and drawing examples on the whiteboard, demonstrated how to use the theorems with the examples of the works, and gave individual assignments to work with in the textbook. In the 45 minutes sessions, the control group students were taught in the whole-class (15 to 20 minutes), and then individual practice (20 to 25 minutes) on the same problems of the textbook with teacher monitoring and guidance. No technology-mediated instruction other than that of a standard classroom (e.g., simple calculators) was allowed. Assignments on homework were the same across groups. The material taught was the same as that of the experimental group, but only the instructional method differed. The qualifications of teachers in both conditions were similar (Bachelor degrees in mathematics education) and the teaching experience was five to eight years.

Fidelity of implementation.

Structured observation protocols were used during the period of intervention to guarantee fidelity of implementation. Unannounced observations were conducted during 20% of randomly selected sessions (three per group) by a trained research assistant through a standardised checklist and in agreement with the planned instructional activities. In the case of the experimental group, it was observed that AI-based tutoring system worked as planned, students used the adaptive features (feedback, hints, visualisations) and teachers gave technical support without instructing the content. In the control group, it was confirmed that the instruction was delivered in the traditional teacher-centered style, and without the introduction of AI tools or other technology-intermediated factors. The internal validity of the study was supported by the fact that all observed sessions were in line with the planned intervention protocols.

3.4 Instruments

Two researcher-developed instruments were used to measure the dependent variables.

i. Geometry Conceptual Understanding Test (GCUT): The GCUT assessed students' understanding of core geometric concepts and relationships. The test consisted of 25 multiple-choice and short-answer items requiring students to interpret diagrams, identify geometric relationships, apply principles to novel situations, and explain reasoning. Items emphasised conceptual understanding rather than procedural computation. Content validity was assessed using the content validity index (CVI). Three mathematics educators rated each item for relevance on a 4-point scale. The item-level CVI (I-CVI) ranged from 0.83 to 1.00, and the scale-level CVI (S-CVI/Ave) was 0.94, exceeding the recommended threshold of 0.90 (Polit & Beck, 2006). A sample GCUT item is: *"In triangles ABC and DEF, $\angle A = \angle D = 50°$, $\angle B = \angle E = 60°$, and AB = DE = 5 cm. Which congruence theorem justifies that the triangles are congruent? Explain your reasoning."* The test demonstrated acceptable internal consistency, with Cronbach's $\alpha = 0.86$ for the pretest and $\alpha = 0.89$ for the posttest.

ii. Geometry Proof Skills Test (GPST): The GPST assessed students' ability to construct and justify geometric proofs. The test comprised five constructed-response proof tasks requiring students to use given conditions, choose applicable theorems and in a logical order, organize the statements, and justify every step. Tasks included two-column proofs and paragraph proofs, which were aimed at formal deductive logic. A sample GPST item presents a diagram with two intersecting lines forming

Social Informatics Journal
Business, Politics, Law, Environmental Sciences & Technology
Vol 12, No 2, 2026 Series
Publication of the Faculty of Computational Sciences, West Midlands Open University, Lagos, Nigeria
Special Issue in collaboration with the International Institute for Multidisciplinary and Development Research

vertical angles and asks: *"Prove that vertical angles are congruent using a two-column proof. Justify each step."* Content validity was checked by experts and S-CVI/Ave was 0.92. The internal consistency was satisfactory, with Cronbach's $\alpha = 0.82$ (pre-test) and $\alpha = 0.85$ (post-test).

Both instruments were piloted with 30 students from a non-participating school. Item difficulty ranged from 0.35 to 0.75, and item discrimination indices exceeded 0.30 for all retained items.

3.5 Data Analysis

Pretest and posttest data were analysed using both descriptive and inferential statistical methods. Descriptive statistics, including means and standard deviations, were computed for each group on both instruments. For inferential analysis, independent-samples t-tests were used to compare posttest scores between the experimental and control groups. This test was appropriate for comparing the means of two independent groups on a continuous outcome variable. Prior to conducting the t-tests, assumptions of normality and homogeneity of variances were examined. No missing data were present in the final sample. Outliers were examined using boxplots and z-scores (threshold $|z| > 3.29$); none were identified. The Shapiro–Wilk test yielded non-significant results for all groups and measures ($p > .05$), indicating that the normality assumption was met. Levene's test was non-significant for both GCUT ($F(1, 62) = 1.24$, $p = .27$) and GPST ($F(1, 62) = 0.89$, $p = .35$), confirming homogeneity of variances. Effect sizes (Cohen's d) and their 95% confidence intervals were calculated to assess the practical magnitude of any observed differences, supplementing statistical significance with an indication of the substantive importance of the results. An alpha level of .05 was used for all statistical tests.

3.6 Ethical Considerations

Ethical approval was obtained from the Institutional Review Board of the authors' institution. Prior to data collection, written informed consent was obtained from all participating students and their parents or guardians. Participants were informed of the study's purpose, the voluntary nature of their participation, and their right to withdraw at any time without consequence. All data were anonymised and stored securely, with access restricted to the research team. No identifying information appears in this manuscript.

4. RESULTS

This section presents the results of the statistical analyses conducted to examine the effect of the AI-based tutoring system on students' conceptual understanding and proof skills in secondary geometry. The analysis is organised around the two research questions.
The results are presented in four stages. First, pretest scores are examined to establish baseline equivalence between the two groups. Second, posttest results pertaining to Research Question 1 are reported. Third, posttest results pertaining to Research Question 2 are presented. Fourth, a summary of the principal findings is provided.

4.1 Pretest Results: Baseline Equivalence

Vol 12, No 2, 2026 Series
Publication of the Faculty of Computational Sciences, West Midlands Open University, Lagos, Nigeria
Special Issue in collaboration with the International Institute for Multidisciplinary and Development Research

Prior to the intervention, both groups completed the GCUT and GPST. Independent-samples t-tests were conducted to determine whether the two groups differed significantly at baseline. Table 1 presents the pretest results.

Table 1
Pretest comparison of experimental and control groups

Variable	Group	N	Mean	Standard Deviation	t-value	p-value
GCUT (Conceptual Understanding)	Experimental	32	41.72	7.84	0.42	.67
	Control	32	40.95	8.12		
GPST (Proof Skills)	Experimental	32	38.63	6.91	0.36	.72
	Control	32	38.01	7.02		

$p > .05$ indicates no statistically significant difference between groups.

No statistically significant differences were observed between the two groups on either measure. For conceptual understanding, the experimental group mean was 41.72 (SD = 7.84) and the control group mean was 40.95 (SD = 8.12); the difference was not statistically significant ($t(62) = 0.42$, $p = .67$). The mean difference of 0.77 points (95% CI [−2.89, 4.43]) had a confidence interval crossing zero, confirming the absence of a meaningful baseline difference. For proof skills, the experimental group mean was 38.63 (SD = 6.91) and the control group mean was 38.01 (SD = 7.02); this difference was also not statistically significant ($t(62) = 0.36$, $p = .72$). The mean difference of 0.62 points (95% CI [−2.81, 4.05]) likewise crossed zero. Both p-values exceeded the .05 significance threshold, confirming that the two groups were comparable prior to the intervention. This baseline equivalence strengthens the interpretation of any posttest differences as being associated with the instructional treatment rather than pre-existing group differences.

4.2 Posttest Results: Research Question 1 — Conceptual Understanding
Is there a significant difference in geometry conceptual understanding between students who receive instruction through an AI-based tutoring system and those who receive conventional teacher-led instruction?

Following the intervention, both groups completed the posttest. An independent-samples t-test was conducted to compare posttest performance on the GCUT. Table 2 presents the results.

Table 2
Posttest comparison for geometry conceptual understanding (GCUT)

Group	N	Mean	SD	t(62)	p-value	Cohen's d	95% CI
Experimental	32	78.46	6.12	9.87	< 0.001	2.33	[1.67, 2.99]
Control	32	61.15	8.03				

The analysis revealed a statistically significant difference in posttest conceptual understanding scores between the two groups ($t(62) = 9.87$, $p < .001$). Students in the experimental group (M =

78.46, SD = 6.12) scored substantially higher than students in the control group (M = 61.15, SD = 8.03), with a mean difference of 17.31 points. The effect size was large (Cohen's d = 2.33, 95% CI [1.67, 2.99]). While this effect size exceeds typical ranges reported in mathematics education meta-analyses (e.g., d ≈ 0.50 - 0.80; Steenbergen-Hu & Cooper, 2014), it is comparable to other short-term ITS studies using researcher-developed instruments (cf. Kestin et al., 2025, who reported d = 1.73 for AI tutoring versus active learning). The magnitude should be interpreted with caution pending replication using standardised assessments and longer follow-up periods. In response to Research Question 1, there was a statistically significant difference in geometry conceptual understanding in favour of students who received AI-based tutoring. Students who were instructed through the AI system demonstrated greater proficiency in interpreting geometric relationships, connecting concepts, and applying geometric principles to novel situations compared with students who received conventional teacher-led instruction.

4.3 Posttest Results: Research Question 2 - Proof Skills

Is there a significant difference in geometry proof skills between students who receive instruction through an AI-based tutoring system and those who receive conventional teacher-led instruction?

An independent-samples t-test was conducted to compare posttest performance on the GPST. Table 3 presents the results.

Table 3
Posttest comparison for geometry proof skills (GPST)

Group	N	Mean	SD	t(62)	p-value	Cohen’s d	95% CI
Experimental	32	74.91	7.03	8.44	< 0.001	1.96	[1.33, 2.59]
Control	32	59.27	8.45				

The analysis revealed a statistically significant difference in posttest proof skills scores between the two groups ($t(62) = 8.44$, $p < .001$). Students in the experimental group (M = 74.91, SD = 7.03) scored substantially higher than students in the control group (M = 59.27, SD = 8.45), with a mean difference of 15.64 points. The effect size was large (Cohen's d = 1.96, 95% CI [1.33, 2.59]). As with the GCUT results, this effect size falls at the upper end of the range reported in the ITS literature (Steenbergen-Hu & Cooper, 2014) and should be interpreted cautiously pending independent replication.

In response to Research Question 2, there was a statistically significant difference in geometry proof skills in favour of students who received AI-based tutoring. Students who instructed through the AI system demonstrated greater competence in arranging logical statements, selecting appropriate theorems, and providing step-by-step justification in deductive proofs compared with students who received conventional instruction.

4.4 Summary of Findings

Overall, the experimental group scored higher than the control group in conceptual understanding (17.31 points) and proof skills (15.64) which were statistically significantly different ($p < .001$) with large effect sizes (Cohen's $d = 2.33$ and 1.96, respectively). These results repeatedly suggest that students instructed with AI had much greater learning outcomes on both of these measures than did students tutored in conventional classrooms. Adaptive characteristics of the AI tutoring system, such as real-time feedback, guided hints, and interactive visualisations, are aligned with the fact that the system facilitated the transition of students from solving procedural problems to constructing rigorous logical proofs, which cannot be directly attributed to any particular features of the system because of the nature of the study design.

5. DISCUSSION

This study examined whether an AI-based tutoring system was associated with improvement in conceptual knowledge and proof skills, of secondary school students in geometry. The findings are encouraging: students who were instructed with the AI-assisted system achieved significantly better results on both outcome measures compared to the students who received the conventional teacher-centred instruction. The improvements were found in both theoretical knowledge and construction of proofs, which indicates that under the circumstances of this experiment, the AI-based environment was related to higher learning gains compared to conventional teaching. A possible answer to these gains is the adaptive response to the tutoring system.

In traditional classrooms feedback would tend to be delayed due to the need to deal with several students at a time. It is possible that errors are not corrected and the student may continue to reason in a faulty manner. This dynamic was changed by the AI tutoring system, which gave feedback in the process of problem-solving, allowing students to identify and rectify mistakes before they proceed to the next steps. This is especially consequential in geometry, where one false premise can nullify a complete line of argument. These results agreed with those of Marwiang et al. (2025), who found that real time feedback in intelligent tutoring systems enhances mathematics learning, by enabling the student to correct the misconceptions in the task, but direct evidence of the particular contribution of the feedback mechanism would need additional experimental isolation.

The interactive visualisation tools in the AI system could also have been a contributor to the gains realised. However, the present design does not permit causal attribution of gains to specific features. Future studies should employ factorial or component-control designs. Geometry relies on seeing connections between lines, angles and forms. Students tend to have difficulty relating diagrammatic representations to formal statements. This gap can be addressed through interactive visualisation, which enables students to interact with figures and see how their properties vary or do not change. In this study, students were able to vary lengths of sides and angles of dynamic triangle figures and could see under what circumstances similarity was maintained, which might be more effective than textbook drawings in helping students to build up the concept of similarity. Such results are compatible with Canonigo (2024) who established that interactive-based AI-assisted learning environments enhanced conceptual knowledge and self-confidence in mathematics.
This enhancement in the levels of proof is in line with the scaffolding that is integrated into the AI system. The cognitive load of proof construction is a challenge as students have to concurrently

locate the given information, identify the purpose of the proof, choose appropriate theorems, arrange statements in logical order and support them. When trying to co-ordinate these multiple demands many students lose metacognitive control. The tutoring system was created to help lessen the cognitive burden by providing students with steps in between, and hints in case of stalling of progress. Such results are in line with other studies on intelligent tutoring systems in mathematics. Lin et al. (2023) indicated that AI-assisted systems enhance the learning of mathematics by offering personalisation and instructional pathways. Yi et al. (2024) also discovered quantifiable benefits in K12 contexts, especially in systems that integrated interactive problem-solving with real-time feedback. The current research builds upon this body of evidence to geometric reasoning and proof-areas that have not gotten as much empirical research as general mathematics achievement.

How do these effect sizes compare with previous ITS studies? Meta-analysis by Steenbergen-Hu and Cooper (2014) indicated an average effect of d = 0.50 of ITS versus conventional teaching in mathematics at K-12, with a wide heterogeneity range (ranging between -.30 and 2.10). Recent analyses of AI tutoring in physics (Kestin et al., 2025, d = 1.73) and initial research of the Geometry Proof Tutor fall within the upper range of this range (Wertheimer, 1990). This implies that geometry, where step-by-step reasoning and visual/spatial processing are heavily involved, can be especially susceptible to AI-based scaffolding. This also suggests either that geometry is uniquely amenable to AI scaffolding, or that methodological factors (researcher-developed tests, pretest sensitisation, novelty effects) inflated the estimates. Nonetheless, the current estimates can be considered as upper-bound approximations whose generalisability needs to be demonstrated until they are replicated with the help of standardised assessments and longer follow-up time.

Some other possible explanations to the learning gains observed should be noted. To begin with, the novelty of the technology might have led to the growth in the engagement and motivation of students, a phenomenon often called the novelty effect. A novelty effect cannot be ruled out, as experimental group students had no prior exposure to AI tutoring systems, and thus their increased interest could have contributed to their effort and persistence during the intervention. Second, the total amount of instruction time was the same across conditions but the quality of student engagement varied. Students of the experimental groups were actively involved in hands-on problem solving with direct feedback, and control group students were taught by the teacher with textbook-based practice.

This distinction in the type of activity and not the AI system itself could be a contributing factor to the differences. Third, sensitisation during pretest could have affected outcomes: both groups took the same pretest and posttest assessment which might have sensitised experimental group students to pay more attention to geometry concepts during intervention (the testing effect). This possibility should be controlled or alternatively used in future studies by a Solomon four-group design (Campbell & Stanley, 1963). Fourth, researcher bias cannot be fully disregarded because the authors created both GeoAI Tutor and the outcome measures. Although assessments were administered by independent research assistants, the possibility of measurement bias in favour of the experimental condition is also a known limitation.

This work is also a part of the current debate regarding the appropriate role of AI in education. Although certain studies show that AI could facilitate mathematical explanation and reasoning (Wardat et al., 2023), other studies warn that AI may hinder learning when students resort to it to get an answer without thinking independently (Bastani et al., 2025). The current research assists in explaining such a tension: the AI system used was integrated into a systematic instructional framework that facilitated, and not substituted student thinking by providing feedback, hints, and guided assignments. It is not whether AI is utilized, but how it is designed and implemented that is most important. The results are of special importance to geometry teaching, as proof is one of the most difficult topics in secondary school mathematics. The findings indicate that AI-based tutoring could support the process of geometric observation to formal justification, but the mechanisms of such facilitation should be directly examined. These results can be particularly relevant in the environments where teaching materials are limited. In large classes, teachers have minimal ability to offer individualised support, and an AI tutoring system can offer more feedback and guided practice than a single teacher can offer during a lesson, which has been observed by Akintade and Olaore (2025) to be consistent with gains in geometry instruction in Nigerian secondary schools.

6. CONCLUSION

This study investigated whether conceptual understanding and proof skills in geometry among secondary school students could be influenced by an AI-based tutoring system in a quasi-experimental pretest-posttest control group design. The experimental group was compared with the control group. The findings indicate that, under the conditions studied — six weeks of instruction, researcher-developed measures and a sample of 64 Nigerian secondary school students, the AI-supported group showed significantly higher learning outcomes than the conventional instruction group on the two outcome measures, with large effect sizes. The results are also part of the increasing body of empirical research related to AI in mathematics education, as they address the specific area of geometric reasoning and proof, which is less empirically studied than overall achievement outcomes. The differences identified are in line with the hypothesis that the adaptive feedback, interactive visualisation, and structured scaffold can facilitate geometry learning. Nevertheless, the unusually large effect sizes should be replicated with the help of standardised measurement and longitudinal study designs before robust judgements regarding magnitudes can be made.

In relation to education practice, the findings imply that AI-based tutoring systems can be useful as an addition to teaching geometry in secondary school, especially when they are tailored to assist students in their reasoning and not to substitute it. Implementation requires special attention to instructional design, training of teachers and integration with existing curricula to ensure that the technology encourages students' mathematical thinking rather than overreliance on computer-generated instructions. Pending independent replication with standardised measures and longer follow-up, these findings should be viewed as promising.

social Informatics Journal
Business, Politics, Law, Environmental Sciences & Technology
Vol 12, No 2, 2026 Series
Publication of the Faculty of Computational Sciences, West Midlands Open University, Lagos, Nigeria
Special Issue in collaboration with the International Institute for Multidisciplinary and Development Research

7. LIMITATIONS AND FUTURE RESEARCH.

A number of limitations should be noted. To begin with, the sample size (64 students) was selected in two schools within one region in Nigeria, which limits the generalizability of the results to other populations. Future research is advised to use larger and more diverse samples and investigate various regions and educational settings to give more credible and transferable evidence. Second, the intervention covered a narrow range of geometry concepts triangle congruence, triangle similarity, and deductive proof. Even though these are the main areas of interest in secondary geometry, the results might not be applicable to all fields of mathematics. Future studies of AI-based tutoring systems should be conducted over a wider range of mathematical topics and teaching contexts.

Third, the research only measured the results at the end of the intervention and not long-term retention. As a result, it is not known whether the observed gains were sustained over time. Longitudinal designs should be used in future research in order to assess how long the gains in conceptual understanding and proof skills will last after AI support is withdrawn. Fourth, although fidelity checks ensured that the desired protocols were followed in one-fifth of the sessions, fidelity monitoring was not done continuously. Further studies are needed to have systematic fidelity checks in all sessions including screen recordings of AI use, and video-recordings of teaching in control group to facilitate fidelity of implementation and analysis of fidelity and outcome. Fifth, the research failed to control the possible teacher effects or prior student experience with technology. Though the conditions had similar teachers in terms of their qualification, the teachers in both conditions could have been different in terms of their teaching style or interest in the intervention. The future studies should employ designs that measure or control these variables.

Sixth, there were no control measures to eliminate the possibility of testing effects, because both groups were tested with the same pretest and posttest measures. Causal inferences in future studies would be enhanced by the Solomon four-group design (Campbell & Stanley, 1963) which separates the effects of a treatment on pretest sensitisation. Lastly, the research was not done to determine if there is any difference in effects by student attributes like previous achievement, mathematics anxiety, or previous technology experience. Future studies are needed to determine whether AI-based tutoring systems are equally effective among various learner profiles and whether certain system characteristics are helpful to a specific subgroup of students. Further research should also explore how AI tutoring systems can be combined with classroom-based instruction in a manner that complements the instructional activities of teachers and retain interest in complex mathematical reasoning, and how adaptive feedback and scaffolding facilitates the construction of proofs, in particular. Fidelity checks were conducted for only 20% of sessions; continuous monitoring (e.g., screen recording) was not employed.

REFERENCES

1. Akintade, C. A., & Olaore, F. A. (2025). Effect of integrating artificial intelligence (AI) on the teaching and learning of geometry within secondary schools in Ogun Central Senatorial Districts, Nigeria. *International Journal of Research and Innovation in Social Science, 9*(9), 1–12. https://doi.org/10.47772/ijriss.2025.909000018
2. Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences of the United States of America, 122*(10), e2422633122. https://doi.org/10.1073/pnas.2422633122
3. Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand McNally.
4. Canonigo, A. (2024). Leveraging AI to enhance students' conceptual understanding and confidence in mathematics. *Journal of Computer Assisted Learning, 40*(6), 3215–3229. https://doi.org/10.1111/jcal.13065
5. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
6. Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4*(4), 253–278.
7. Duval, R. (1998). Geometry from a cognitive point of view. In C. Mammana & V. Villani (Eds.), *Perspectives on the teaching of geometry for the 21st century* (pp. 37–52). Kluwer.
8. Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149
9. Fitri, P., Hartono, Y., & Meryansumayeka, M. (2025). Learning proof of trigonometric identities with ChatGPT. *Journal of Honai Math, 8*(1), 1–14. https://doi.org/10.30862/jhm.v8i1.755
10. Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 805–842). Information Age Publishing.
11. Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports, 15*, Article 8573. https://doi.org/10.1038/s41598-025-97652-6
12. Lin, C., Huang, A., & Lu, O. (2023). Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review. *Smart Learning Environments, 10*, Article 41. https://doi.org/10.1186/s40561-023-00260-y
13. Marwiang, M., Prasertsang, M., & Junpeng, P. (2025). Enhancing students' learning outcomes in mathematics through intelligent tutoring systems based on real-time feedback. *Journal of Education and Learning, 14*(6), 186–197. https://doi.org/10.5539/jel.v14n6p186
14. Menlah, C., & Boateng, F. (2025). Examining the effect of AI-based tutoring systems on students' mathematical problem-solving skills: The moderating role of mathematics anxiety. *Journal of Pedagogical Sociology and Psychology, 7*(2), 112–125. https://doi.org/10.33902/jpsp.202536137

15. Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
16. Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology, 106*(2), 331–347. https://doi.org/10.1037/a0034801
17. Stylianides, G. J., & Stylianides, A. J. (2009). Facilitating the transition from empirical arguments to proof. *Journal for Research in Mathematics Education, 40*(3), 314–352.
18. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
19. Wardat, Y., Tashtoush, M., Alali, R., & Jarrah, A. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education, 19*(7), em2286. https://doi.org/10.29333/ejmste/13272
20. Wertheimer, R. (1990). The Geometry Proof Tutor: An "intelligent" computer-based tutor in the classroom. *The Mathematics Teacher, 83*(4), 308–317.
21. Xing, W., Song, Y., Li, C., Liu, Z., Zhu, W., & Oh, H. (2025). Development of a generative AI-powered teachable agent for middle school mathematics learning: A design-based research study. *British Journal of Educational Technology, 56*(2), 456–475. https://doi.org/10.1111/bjet.13586
22. Yi, L., Liu, D., Jiang, T., & Xian, Y. (2024). The effectiveness of AI on K–12 students' mathematics learning: A systematic review and meta-analysis. *International Journal of Science and Mathematics Education, 23*, 1105–1126. https://doi.org/10.1007/s10763-024-10499-7