PALGO JOURNAL OF EDUCATION RESEARCH
ISSN 2476-8332
DOI : https://doi.org/10.65756/pjer.2026.007
Volume 11 Issue 3, August 2026.Page 39-57
http://www.palgojournals.org/PJER/Index.htm
Corresponding Author's Email:nangolo1987@gmail.com

The Role of Generative Artificial Intelligence in Enhancing Mathematical Reasoning Among Secondary School Learners

Philipus N Nangolo[1,*], Secilia M Nafenya[2], Virginia M Moyo[2], Sebastian M Munenda[2]

Department of Mathematics, International University of Management (IUM), Nkurenkuru Campus, Namibia[1]
Department of Education, International University of Management (IUM), Nkurenkuru Campus, Namibia[2]

Accepted ,July 22 , 2026

Generative Artificial Intelligence (GenAI) provides personalised explanations, immediate feedback, and interactive problem-solving support; however, evidence of its effectiveness in strengthening mathematical reasoning in developing educational contexts remains limited. This study developed an AI-Assisted Mathematics Learning Model and illustrated its evaluation through a quasi-experimental mixed-methods design involving 120 secondary school learners and 10 mathematics teachers. Learners were divided equally between a GenAI-assisted experimental group and a conventional-instruction control group. Data were obtained through mathematical reasoning tests, questionnaires, classroom observations, and semi-structured interviews. In the analysis, the experimental group achieved a significantly higher adjusted post-test mean than the control group, $F(1,117) = 180.96$, $p < .001$, partial $\eta^2 = .607$. Synthetic qualitative evidence illustrated potential benefits related to personalised explanations, immediate feedback, strategy comparison, and reflective problem-solving, while highlighting concerns about inaccurate responses, learner dependence, technological access, and teacher mediation. The proposed model provides a structured framework for investigating the responsible integration of GenAI into Namibian secondary school mathematics education.

Keywords: Generative Artificial Intelligence; mathematics education; mathematical reasoning; secondary school learners.

1. Introduction

1.1 Background of the Study

Mathematical reasoning is a central objective of mathematics education because it enables learners to analyse problems, formulate conjectures, construct logical arguments, justify solutions, and apply mathematical knowledge in routine and non-routine situations (Kilpatrick et al., 2001 ; National Council of Teachers of Mathematics, 2014). Accordingly, contemporary mathematics curricula increasingly emphasise conceptual understanding, critical thinking, problem-solving, and mathematical communication rather than procedural fluency alone. Nevertheless, many secondary school learners continue to experience difficulty in developing these competencies, particularly in learning environments dominated by memorisation, algorithmic procedures, and teacher-centred instruction.

Recent developments in artificial intelligence (AI), especially generative artificial intelligence (GenAI), have created new opportunities for supporting mathematics teaching and learning. GenAI applications such as ChatGPT, Gemini, and Microsoft Copilot can generate personalised explanations, provide immediate feedback, suggest alternative solution strategies, and facilitate interactive mathematical discussions. These capabilities may strengthen conceptual understanding and mathematical reasoning by allowing learners to examine multiple representations, compare solution methods, identify errors, and receive support throughout the problem-solving process (Kasneci et al., 2023; Miao and Holmes, 2023; Walkington, 2025).

The educational value of GenAI, however, depends on how it is incorporated into classroom practice. Uncritical dependence on AI-generated solutions may reduce learners' engagement in independent reasoning and encourage the reproduction of answers without adequate conceptual understanding. GenAI systems may also produce inaccurate, incomplete, or misleading mathematical explanations. Teacher mediation, verification of AI-generated content, and explicit instruction in the critical use of AI are therefore necessary for meaningful and responsible implementation (Giannakos et al., 2025; Kasneci et al., 2023; Miao and Holmes, 2023; UNESCO, 2021b).

Evidence from Namibia has begun to illustrate both the possibilities and constraints associated with AI-supported mathematics education. For example, Elifas and Simuja (2024) examined the experiences of 13 mathematics teachers who integrated AI technologies into rural upper-primary mathematics classrooms in Namibia. The study reported perceived improvements in learners' mathematical comprehension, visualisation, and problem-solving. However, it also identified inadequate infrastructure, limited access to digital devices, insufficient teacher training, and a shortage of culturally relevant AI resources. Similarly, Stefanus et al. (2025) found that Namibian teachers demonstrated positive intentions to adopt AI, although limited resources and low perceived behavioural control remained important implementation barriers.

These studies provide valuable contextual evidence, but they do not establish whether GenAIassisted instruction produces measurable improvements in the mathematical reasoning of secondary school learners. In particular, limited quasi-experimental evidence is available comparing GenAI-assisted instruction with conventional teaching while simultaneously examining learners' and teachers' experiences. The present study addressed this gap by investigating the effect of structured GenAI-assisted instruction on learners' logical reasoning, conceptual understanding, problem-solving, and mathematical communication. It also developed evidence-based recommendations for the responsible integration of GenAI into Namibian secondary school mathematics classrooms.

1.2 Problem Statement

Although GenAI can provide personalised explanations, immediate feedback, and interactive problem-solving assistance, its availability does not necessarily result in improved mathematical reasoning. Learners may use these systems primarily to obtain final answers rather than to construct arguments, evaluate strategies, justify procedures, or reflect critically on mathematical relationships. Consequently, poorly structured GenAI use may reinforce superficial learning and technological dependence instead of strengthening independent reasoning (Kasneci et al., 2023 ; Miao and Holmes, 2023; Giannakos et al., 2025).

This concern is particularly relevant in Namibia, where AI-supported education is developing within a context characterised by unequal access to digital infrastructure, limited teacher preparation, and insufficient locally grounded evidence. Existing Namibian research has mainly examined teachers' experiences, perceptions, and intentions regarding the adoption of AI (Elifas and Simuja, 2024; Stefanus et al., 2025). It has not adequately determined whether structured GenAI-assisted instruction improves secondary school learners' mathematical reasoning relative to conventional teaching methods.

The central problem addressed by this study was therefore the absence of context-specific empirical evidence concerning the effectiveness of GenAI-assisted instruction in developing mathematical reasoning among Namibian secondary school learners. Without such evidence, teachers, curriculum developers, and policymakers have an insufficient basis for deciding how GenAI should be incorporated into mathematics classrooms. This study consequently compared the mathematical reasoning performance of learners exposed to structured GenAI-assisted instruction with that of learners taught through conventional methods. It also examined learners' and mathematics teachers' perceptions of the educational value, limitations, and responsible use of GenAI.

2. Literature Review

2.1 Teaching Mathematical Reasoning

Mathematical reasoning refers to the process through which learners analyse mathematical relationships, formulate conjectures, construct and evaluate arguments, justify procedures, and draw logically defensible conclusions. It extends beyond the accurate execution of algorithms because it requires learners to explain why a procedure is valid, identify the assumptions underlying a solution, and transfer mathematical knowledge to unfamiliar situations (Kilpatrick et al., 2001; National Council of Teachers of Mathematics, 2014; Schoenfeld, 1985). Mathematical reasoning therefore encompasses inductive reasoning, through which patterns are used to formulate conjectures; deductive reasoning, through which conclusions are derived from accepted premises; and analogical reasoning, through which structural similarities are used to connect mathematical situations.

Contemporary mathematics curricula regard reasoning, problem-solving, communication, and conceptual understanding as mutually reinforcing competencies. Learners who can perform procedures but cannot justify them may demonstrate instrumental knowledge without developing a relational understanding of the mathematical concepts involved (Skemp, 1976). Effective mathematics instruction should therefore create opportunities for learners to pose questions, compare strategies, explain relationships, identify errors, defend conclusions, and revise their thinking in response to evidence.

The development of mathematical reasoning depends substantially on the nature of classroom instruction. Teacher-centred practices that emphasise memorisation, examination preparation, and the reproduction of standard procedures may improve short-term procedural performance but provide limited opportunities for conjecturing, justification, and mathematical discourse. In contrast, instruction that incorporates inquiry, worked-example comparison, problem variation, collaborative discussion, and reflective questioning can support deeper conceptual understanding (Schoenfeld, 1985; Polya, 1957; Hattie, 2009). Digital technologies may extend these opportunities when they are used to stimulate reasoning rather than merely accelerate answer production (Drijvers, 2023; Hoyles and Lagrange, 2010).

2.2 Artificial Intelligence and Generative AI in Mathematics Education

Artificial intelligence in education encompasses computational systems that perform tasks associated with human cognition, including prediction, classification, natural-language interaction, adaptive feedback, and decision support

(Russell and Norvig, 2021; Holmes et al., 2019). Earlier educational AI systems commonly operated through predefined knowledge models, rule-based tutoring, or adaptive sequences. Generative artificial intelligence differs from these systems because it can generate new text, explanations, worked examples, questions, computer code, and mathematical representations in response to natural-language prompts (Kasneci et al., 2023 ; Miao and Holmes, 2023).

Large language models underlying applications such as ChatGPT, Gemini, Claude, and Microsoft Copilot allow learners to engage in conversational exchanges about mathematical ideas. A learner can request an explanation, question a particular step, compare alternative solution methods, ask for a simpler example, or request feedback on an attempted solution. These affordances make GenAI potentially valuable as an interactive form of scaffolding rather than only as an answer-generating mechanism (Kasneci et al., 2023; Walkington, 2025).

In mathematics education, GenAI may support learning through at least four mechanisms. First, it can provide immediate explanatory feedback when teacher support is not instantly available. Second, it can generate multiple representations and solution strategies, thereby helping learners recognise connections among symbolic, graphical, numerical, and verbal forms. Third, it can adapt explanations to different levels of prior knowledge. Fourth, it can prompt learners to explain, critique, and revise their reasoning (Walkington, 2025; Biton et al., 2025; Wardat et al., 2023). These functions are pedagogically valuable when learners remain cognitively engaged in constructing and evaluating solutions.

The same capabilities also introduce important limitations. Large language models generate responses by estimating plausible linguistic sequences rather than by applying guaranteed systems of mathematical verification. They may therefore produce correct-looking but mathematically invalid solutions, incomplete proofs, inconsistent notation, or explanations that conceal computational errors (Kasneci et al., 2023; Wardat et al., 2023). The quality of the output may also depend on the wording of the prompt, the complexity of the problem, and the learner's ability to recognise an erroneous response.

Consequently, GenAI does not automatically improve mathematical reasoning. When learners use it principally to obtain completed solutions, it may reduce productive struggle, independent problem-solving, and metacognitive monitoring. When used within carefully designed activities that require prediction, justification, comparison, error analysis, and reflection, it may instead serve as a cognitive and dialogical scaffold. The pedagogical effect therefore depends not merely on access to the technology but on task design, teacher guidance, learners' AI literacy, and the level of intellectual engagement required (Miao and Holmes, 2023; UNESCO, 2021 b; Giannakos et al., 2025).

2.3 Pedagogical and Ethical Conditions for Effective Integration

Effective GenAI integration requires a clear distinction between assistance and substitution. Assistance occurs when AI-generated content helps learners interpret a concept, examine an alternative strategy, locate an error, or receive feedback on their own reasoning. Substitution occurs when the system performs the central reasoning task on behalf of the learner. The former may strengthen learning, whereas the latter may create an illusion of competence without corresponding conceptual development.

Teacher guidance is therefore central to GenAI-assisted instruction. Teachers must design prompts and activities that require learners to evaluate generated solutions, provide mathematical justification, compare AI-generated procedures with their own reasoning, and identify possible errors. They must also ensure that learners understand that fluency and confidence in an AI-generated response do not establish mathematical validity (Miao and Holmes, 2023 ; UNESCO, 2021b). In this sense, GenAI literacy in mathematics includes not only operational knowledge of the tool but also the ability to verify calculations, evaluate assumptions, recognise hallucinations, and judge whether a response addresses the original problem.

Ethical considerations include privacy, data protection, algorithmic bias, academic integrity, unequal technological access, and transparency regarding AI use. These concerns are particularly important in resource-constrained educational contexts, where differences in internet connectivity, device availability, language, and digital literacy may produce unequal learning opportunities. Responsible integration therefore requires appropriate school policies, teacher preparation, equitable access, and assessment practices that continue to evaluate learners' independent reasoning (Miao and Holmes, 2023; UNESCO, 2021b; Namibia Ministry of Education, Arts and Culture, 2023).

2.4 Theoretical Foundations

The study was informed by Constructivist Learning Theory, Social Constructivism, Cognitive Load Theory, Metacognitive Theory, and the Technology Acceptance Model. These perspectives explain different but complementary dimensions of GenAI-assisted mathematics learning.

Constructivist Learning Theory holds that learners actively construct knowledge by connecting new experiences with existing cognitive structures (Piaget, 1950; Bruner, 1960). From this perspective, learning does not occur simply because GenAI presents an explanation. It occurs when learners interpret the explanation, compare it with prior knowledge, test its validity, and reorganise their understanding. GenAI-assisted instruction is therefore consistent with constructivism only when learners actively investigate and evaluate mathematical ideas rather than passively reproduce generated answers.

Social Constructivism emphasises the role of interaction, language, and scaffolding in learning. Within Vygotsky's Zone of Proximal Development, learners can accomplish more complex tasks when they receive support that is responsive to their current level of understanding (Vygotsky, 1978). GenAI may provide supplementary scaffolding through questions, hints, examples, and feedback. However, it cannot fully replace the teacher's pedagogical judgement, awareness of classroom context, or ability to interpret learners' emotional and conceptual needs.

Cognitive Load Theory distinguishes intrinsic cognitive load, which arises from the complexity of the learning content; extraneous load, which results from ineffective presentation; and germane cognitive activity directed towards schema construction (Sweller, 2020; Sweller et al., 2011). GenAI may reduce extraneous load by clarifying notation, dividing complex procedures into manageable steps, and providing appropriately sequenced examples. Conversely, overly detailed, unstructured, or inaccurate responses may increase cognitive load and distract learners from the underlying mathematical relationships. The theory therefore supports the use of concise, structured, and learner-appropriate AI feedback.

Metacognitive Theory concerns learners' capacity to plan, monitor, and evaluate their thinking (Flavell, 1979). GenAI-assisted activities may strengthen metacognition when learners are required to predict an answer before consulting the system, explain why a generated procedure is correct, identify doubtful steps, compare strategies, and reflect on changes in their understanding. Such activities position learners as evaluators of AI-generated content rather than passive recipients.

The Technology Acceptance Model proposes that technology adoption is substantially influenced by perceived usefulness and perceived ease of use (Davis, 1989; Venkatesh and Davis, 2000). Learners and teachers may be more willing to use GenAI when they believe that it improves explanation, feedback, lesson preparation, or problem-solving efficiency. Acceptance alone, however, does not demonstrate educational effectiveness. A tool may be perceived as useful because it produces answers quickly even when its use does not strengthen reasoning. Accordingly, this study considered both participants' perceptions and measured mathematical reasoning outcomes.

2.5 Empirical Evidence on GenAI and Mathematics Learning

The international empirical literature generally indicates that GenAI can improve mathematics learning, but the magnitude and nature of its effects vary across instructional contexts. In a meta-analysis of 22 empirical studies, comprising 46 independent samples and 5,232 participants, Liu et al. (2026) reported a moderate positive overall effect of GenAI on mathematics learning outcomes, $g = 0.534$, $p < 0.001$. The effect on cognitive outcomes was $g = 0.596$, while the effect on higher-order cognitive skills was larger, $g = 0.718$. These results suggest that GenAI can support more than procedural performance when it is incorporated into tasks requiring higher-order engagement.

The same meta-analysis, however, demonstrated that positive effects were not uniform. For secondary school learners, the estimated effect was comparatively smaller, $g = 0.313$, although it remained statistically significant. The effect on non-cognitive outcomes was $g = 0.299$ and did not reach conventional statistical significance, $p = 0.052$. Effects also differed according to mathematical content, sample size, learning mode, and the level of GenAI integration. Studies using smaller samples produced larger estimates than larger studies, raising the possibility of small-sample inflation. These findings indicate that positive overall results should not be interpreted as evidence that every form of GenAI use is effective (Liu et al., 2026).

Classroom and perception studies similarly report a mixture of benefits and limitations. Wardat et al. (2023) found that ChatGPT could generate explanations and support mathematical problem-solving, but its responses were not uniformly accurate and required critical verification. Its conversational format enabled users to request clarification and alternative methods, yet plausible presentation sometimes made errors difficult for inexperienced learners to detect. This finding reinforces the importance of mathematical verification and teacher oversight.

In a mixed-methods study involving 125 secondary school learners from two schools in Nigeria, Egara et al. (2025) found that learners valued ChatGPT for supporting understanding, problem-solving, immediate assistance, and personalised learning. Learners nevertheless expressed concerns about its limitations, consistent use, and reliability. Although the study provided important evidence from an African secondary school context, it relied on purposive sampling and primarily investigated perceptions. It did not employ an experimental comparison to determine whether ChatGPT caused measurable improvement in mathematical reasoning.

In South Africa, Motseki and Jojo (2025) conducted interviews and lesson observations with 10 Grade 12 mathematics teachers using ChatGPT in geometry instruction. Teachers perceived ChatGPT as useful for providing basic geometrical explanations and instructional support. However, the system was less effective in correcting misconceptions, and the accuracy of its solutions depended on the cognitive demand of the question, the information supplied, and the formulation of the prompt. The study demonstrated the importance of teacher judgement but did not measure changes in learners' mathematical reasoning or compare ChatGPT-assisted instruction with conventional teaching.

Taken together, the international and African findings suggest that GenAI has pedagogical potential, but much of the available literature remains based on perceptions, exploratory cases, short interventions, or general achievement measures. Mathematical achievement and mathematical reasoning are related but not identical outcomes. A learner may obtain a higher score through procedural assistance without demonstrating stronger capacity to formulate conjectures, construct arguments, justify procedures, or communicate mathematical ideas. Evidence focusing specifically on these reasoning dimensions remains limited.

2.6 Empirical Evidence from Namibia

Namibian research on AI in education is emerging but remains concentrated on teacher experiences, adoption intentions, and implementation conditions. Elifas and Simuja (2024) conducted a qualitative intervention study involving 13 mathematics teachers in rural upper-primary schools. The teachers used technologies such as GeoGebra, Khan Academy, virtual tutors, Photomath, Mathway, and Microsoft Math Solver. They reported that these tools supported visualisation, learner-paced instruction, comprehension, and problem-solving. The study also identified unreliable infrastructure, limited device access, insufficient teacher preparation, and a shortage of culturally relevant AI resources as significant barriers.

Although the study by Elifas and Simuja (2024) provided important contextual evidence, its findings were based mainly on teachers' pedagogical experiences in upper-primary schools and included several forms of educational technology rather than focusing exclusively on GenAI. It did not employ a control group, measure secondary school learners' mathematical reasoning, or estimate the effect of GenAI-assisted instruction.

Using survey data from 159 Namibian in-service teachers, Jatileni et al. (2024) examined factors influencing teachers' behavioural intentions to teach AI. Structural equation modelling showed that perceived AI relevance, attitudes towards AI, confidence, and perceptions of AI for social good contributed to behavioural intention. AI anxiety and readiness were not significant predictors. The study established that teacher acceptance depends on more than access alone, but it examined intentions to teach AI rather than the use of GenAI to teach mathematics or its effect on learner outcomes.

Similarly, Stefanus et al. (2025) examined AI adoption among 22 teachers in the Khomas Region. Participants demonstrated positive attitudes and strong intentions to teach AI, but limited resources reduced their perceived behavioural control. The small sample restricted predictive analysis, and the study did not assess classroom learning outcomes. Collectively, these Namibian studies show growing interest in AI-supported education while also identifying infrastructure, training, access, and contextual relevance as persistent challenges.

The available Namibian evidence therefore establishes the feasibility and perceived value of AI integration but does not demonstrate whether structured GenAI-assisted mathematics instruction improves secondary school learners' reasoning. It also provides little evidence concerning how learners themselves experience GenAI-supported mathematics instruction or how learner and teacher perspectives correspond with objectively measured outcomes.

2.7 Research Gap

The reviewed literature revealed five interrelated gaps. First, there is an *outcome gap*. Much of the literature examines general achievement, technology acceptance, engagement, or perceived usefulness rather than mathematical reasoning as a multidimensional construct comprising logical reasoning, conceptual understanding, problem-solving, justification, and mathematical communication.

Second, there is a *methodological gap*. A substantial proportion of African and Namibian studies use descriptive surveys, qualitative cases, or perception-based designs. Such studies explain how teachers and learners view AI, but they cannot determine whether GenAI-assisted instruction causes improvement relative to conventional teaching. Rigorous quasi-experimental evidence from authentic secondary school classrooms remains limited.

Third, there is a *population gap*. Existing Namibian research has focused mainly on teachers, in-service teacher adoption, or upper-primary settings (Elifas and Simuja, 2024; Jatileni et al., 2024; Stefanus et al., 2025). Secondary school learners' direct experiences and measured mathematical reasoning outcomes have received insufficient attention.

Fourth, there is an *intervention gap*. Previous studies frequently consider AI tools collectively or examine unrestricted access to ChatGPT. There is limited evidence concerning structured GenAI-assisted instruction in which teacher guidance, prompt design, verification, reflection, and independent problem-solving are deliberately incorporated into the intervention.

Fifth, there is a *contextual gap*. International evidence indicates that GenAI effects depend on sample size, mathematical content, instructional design, and the degree of technology integration (Liu et al., 2026). Findings from resource-rich settings cannot therefore be transferred automatically to Namibian schools, where infrastructure, device access, teacher preparation, language, and digital literacy may influence implementation.

This study addressed these gaps through a quasi-experimental mixed-methods design comparing structured GenAI-assisted mathematics instruction with conventional instruction among Namibian secondary school learners. It measured changes in mathematical reasoning, examined the experiences of learners and teachers, and considered the pedagogical and contextual conditions influencing implementation. The study consequently extended existing research from technology acceptance and perceived usefulness towards evidence concerning instructional effectiveness, reasoning outcomes, and responsible classroom integration.

2.8 Conceptual Framework

The conceptual framework proposes that structured GenAI-assisted instruction, the independent variable, influences mathematical reasoning, the dependent variable. Mathematical reasoning is represented through logical reasoning, conceptual understanding, problem-solving, justification, and mathematical communication.

The relationship is explained through four pedagogical mechanisms: learner engagement, cognitive scaffolding, metacognitive reflection, and self-regulated learning. GenAI may increase engagement through interactive dialogue; provide cognitive scaffolding through hints, examples, and feedback; promote metacognition through error analysis and strategy comparison; and support self-regulation by allowing learners to seek assistance and monitor their progress. These mechanisms are treated as explanatory processes rather than statistically tested mediating variables unless a formal mediation analysis is conducted.

The effectiveness of GenAI-assisted instruction is conditioned by teacher guidance, the accuracy of AI-generated content, learner and teacher acceptance, digital access, AI literacy, task design, and ethical use. Teacher guidance is especially important because it determines whether GenAI supports learners' reasoning or substitutes for it. The framework therefore positions GenAI as a complementary instructional resource embedded within teacher-directed pedagogy, rather than as an autonomous replacement for mathematics instruction.

Figure 1 presents the proposed relationships.

3. Methodology

3.1 Research Paradigm and Design

This study was grounded in the pragmatist research paradigm, which permits the integration of quantitative and qualitative methods when addressing a complex educational problem. Pragmatism was appropriate because the study examined both the measurable effect of Generative Artificial Intelligence (GenAI)-assisted instruction on mathematical reasoning and the experiences and perceptions of the learners and teachers who participated in the intervention (Creswell and Creswell, 2023; Johnson and Onwuegbuzie, 2004).

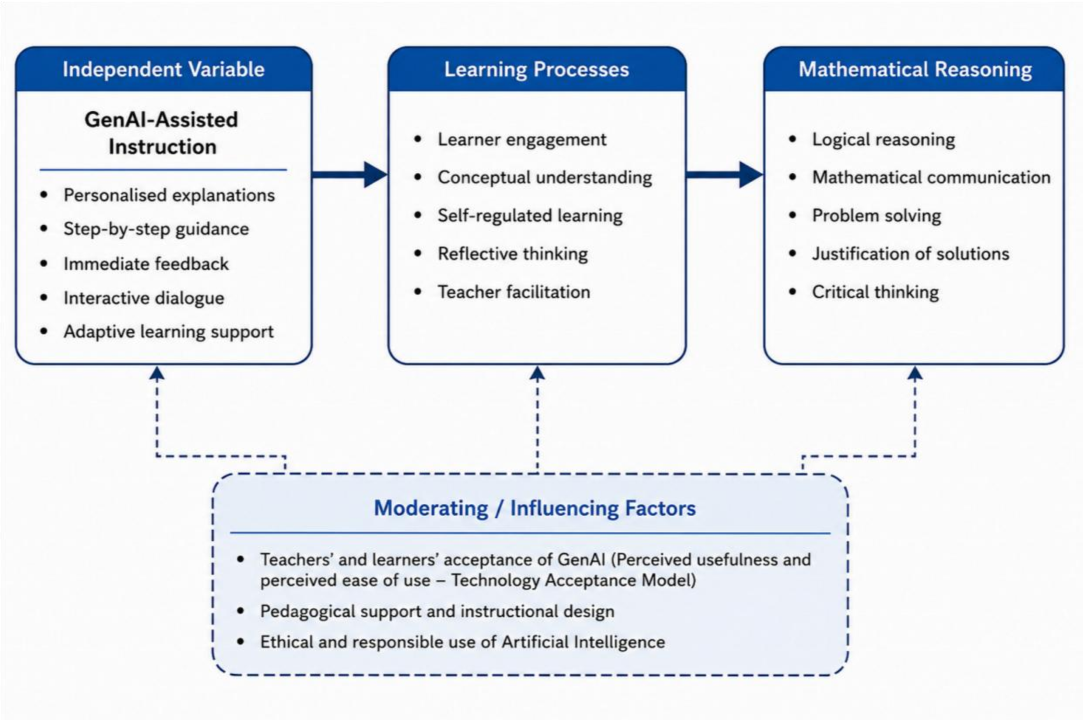

Figure 1: Conceptual framework linking structured GenAI-assisted instruction to secondary school learners' mathematical reasoning.

A quasi-experimental mixed-methods design was employed. The quantitative component followed a pre-test–post-test non-equivalent control group design, while the qualitative component involved learner questionnaires, classroom observations, and semi-structured teacher interviews. The quantitative and qualitative evidence was integrated during interpretation to provide a more comprehensive explanation of the effectiveness, implementation, and perceived value of GenAI-assisted mathematics instruction (Creswell and Creswell, 2023; Campbell and Stanley, 1963; Creswell and Plano Clark, 2024).

Individual randomisation was not feasible because the learners were already organised into established classes, and reorganising them would have disrupted school timetables, normal instructional arrangements, and curriculum delivery. Intact classes were therefore used to preserve the natural classroom setting and minimise contamination between the experimental and control conditions. Because the resulting groups could not be assumed to be equivalent at baseline, pre-test scores were collected and subsequently treated as a covariate in the ANOVA model.

3.2 Study Setting, Population, and Sampling

The study was conducted in selected secondary schools in Namibia. A multistage non-probability sampling procedure was used. First, the participating schools were purposively selected because they possessed the minimum Information and Communication Technology infrastructure required for the intervention, including internet connectivity and access to suitable digital devices. Schools also had to offer the relevant secondary school mathematics curriculum and provide administrative permission for the study.

Second, intact mathematics classes within the selected schools were used as naturally occurring clusters. Eligible learners in the selected classes were invited to participate, subject to informed consent and parental or guardian consent for learners younger than 18 years. This procedure produced a sample of 120 secondary school learners. Sixty learners constituted the experimental group, while 60 learners constituted the non-equivalent control group.

Ten mathematics teachers who taught the participating classes were purposively recruited based on their involvement in the relevant mathematics curriculum, availability during the intervention, and willingness to participate in the interviews and classroom implementation. The purposive selection of schools and teachers was necessary because the intervention required appropriate technological infrastructure and teachers who could support and monitor GenAIassisted learning. Although this sampling procedure limited statistical generalisability, it enabled the study to examine the intervention under realistic classroom conditions.

3.3 Instructional Intervention

The intervention was conducted over eight weeks. Learners in the experimental group received mathematics instruction based on the AI-Assisted Mathematics Learning Model (AIAMLM). During the intervention, learners used approved GenAI applications to obtain personalised explanations, examine worked examples, receive immediate feedback, identify errors, and compare alternative solution strategies. Learners were required to verify AI-generated responses, explain the reasoning underlying each solution, and consult the teacher whenever a response was incomplete, inconsistent, or mathematically questionable (Kasneci et al., 2023; Miao and Holmes, 2023; Biton et al., 2025).

The control group received conventional mathematics instruction without structured GenAI assistance. Both groups studied the same curriculum content over the same period and were given comparable instructional time and mathematical tasks. The principal difference between the conditions was the structured use of GenAI in the experimental group. Both groups completed the Mathematical Reasoning Test before and after the intervention.

To promote implementation consistency, participating teachers followed the planned instructional sequence and covered the same specified mathematical content. Classroom observations were used to document teacher adherence, learner participation, GenAI use, reasoning activities, and possible deviations from the intended instructional procedures.

3.4 Data-Collection Instruments

Data were collected using the Mathematical Reasoning Test, a learner questionnaire, a classroom observation schedule, and a semi-structured teacher interview guide. The use of multiple instruments enabled methodological triangulation and provided complementary evidence concerning learning outcomes, participant perceptions, and classroom implementation.

3.4.1 Mathematical Reasoning Test

The Mathematical Reasoning Test was administered as a pre-test and post-test to both groups. It assessed learners' ability to analyse mathematical problems, recognise patterns and relationships, construct logical arguments, justify procedures, solve unfamiliar problems, and communicate mathematical ideas. The same assessment domains and scoring criteria were applied to the experimental and control groups.

The test was reviewed by subject and research-methodology experts to establish content validity. The reviewers examined the relevance, clarity, curriculum alignment, cognitive demand, and coverage of the mathematical reasoning construct. Their recommendations were used to revise ambiguous instructions, improve item wording, and ensure adequate representation of the intended reasoning domains.

The instrument was pilot-tested with learners who had characteristics similar to those of the main sample but who did not participate in the final study. Internal consistency was assessed using Cronbach's alpha. The Mathematical Reasoning Test produced a reliability coefficient of

$\alpha = 0.84$.

A coefficient of at least 0.70 was considered evidence of acceptable internal consistency for the intended research purpose (Nunnally and Bernstein, 1994).

3.4.2 Learner Questionnaire

The learner questionnaire collected information about learners' experiences with GenAI-assisted mathematics instruction, including perceived usefulness, ease of use, engagement, feedback quality, conceptual understanding, and possible challenges. The questionnaire contained closedended Likert-scale items and open-ended questions. Expert review was used to assess the clarity, relevance, and alignment of the items with the research objectives.

3.4.3 Classroom Observation Schedule

The observation schedule was used to document how the intervention was implemented in the experimental classrooms. Particular attention was given to learner engagement, prompt formulation, interaction with GenAI, verification of AI-generated responses, mathematical discussion, teacher mediation, and the identification and correction of misconceptions. The same schedule was used across observation sessions to promote consistency.

3.4.4 Semi-Structured Teacher Interviews

Semi-structured interviews were conducted with the 10 participating mathematics teachers after the intervention. The interviews explored teachers' experiences of implementing GenAI-assisted instruction, perceived changes in learners' mathematical reasoning, instructional benefits, implementation challenges, ethical concerns, and recommendations for future classroom integration. The semi-structured format provided consistency across interviews while allowing participants to elaborate on issues emerging from their classroom experiences.

3.5 Data-Collection Procedure

Permission to conduct the study was obtained from the relevant educational authorities and participating schools. After informed consent had been obtained, the Mathematical Reasoning Test was administered to the experimental and control groups as a pre-test. The eight-week instructional intervention was then implemented.

During the intervention, classroom observations were conducted to monitor instructional delivery and learner participation. At the end of the intervention, both groups completed the post-test. Learners in the experimental group also completed the questionnaire, and the participating mathematics teachers took part in semi-structured interviews. All instruments were administered under comparable conditions to minimise procedural variation.

3.6 Data Analysis

Quantitative data were analysed using IBM SPSS Statistics, Version 29. Frequencies and percentages were used to summarise categorical data, while means and standard deviations were used to describe learners' pre-test and post-test performance and questionnaire responses.

An independent-samples *t*-test was used to examine baseline differences between the experimental and control groups. Paired-samples *t*-tests were used to determine whether each group demonstrated statistically significant change between the pre-test and post-test. Analysis of covariance (ANOVA) was used as the primary between-group analysis, with post-test mathematical reasoning scores as the dependent variable, instructional group as the independent variable, and pre-test scores as the covariate. ANOVA was selected to adjust the post-test comparison for possible baseline differences arising from the use of non-equivalent intact groups.

Before conducting the inferential analyses, the relevant statistical assumptions were examined. These included independence of observations, approximate normality, absence of extreme outliers, homogeneity of variance, linearity between the covariate and outcome, and homogeneity of regression slopes. Statistical significance was evaluated at the 5% level ($\alpha = 0.05$). Cohen's *d* was used to describe the magnitude of mean differences, while partial eta squared (η_{p}^2) was reported for the ANOVA effect. Reporting effect sizes alongside *p*-values enabled the practical importance of the intervention to be evaluated (Field, 2024).

Qualitative data from interviews, open-ended questionnaire responses, and classroom observations were analysed using thematic analysis. The analysis involved familiarisation with the data, initial coding, grouping related codes, developing candidate themes, reviewing and naming the themes, and producing the final interpretation (Braun and Clarke, 2006). Evidence from the different qualitative sources was compared to identify convergence, complementarity, and contradictions. The qualitative findings were subsequently integrated with the quantitative results during interpretation.

3.7 Ethical Considerations

Ethical approval was obtained from the relevant institutional and educational authorities before data collection commenced. Participation was voluntary, and all participants were informed about the purpose, procedures, potential benefits, and possible risks of the study. Written informed consent was obtained from teachers and adult participants, while parental or guardian consent and learner assent were obtained for participants younger than 18 years.

Participants were informed that they could withdraw from the study without penalty. Codes rather than personal names were used during data analysis and reporting to protect confidentiality. Data were stored securely and accessed only for research purposes. Learners were also instructed not to enter personal or sensitive information into GenAI applications. The intervention emphasised responsible AI use, critical verification of generated responses, academic integrity, and the continuing authority of the mathematics teacher (UNESCO, 2021b; World Medical Association, 2013).

3.8 Development of the AI-Assisted Mathematics Learning Model

The AI-Assisted Mathematics Learning Model (AIAMLM) was developed as a structured pedagogical framework for integrating GenAI into secondary school mathematics instruction. The model was based on the assumptions that learners possessed basic digital literacy, teachers facilitated AI-supported learning, GenAI provided guidance without replacing classroom instruction, learners critically evaluated AI-generated responses, and all learning activities complied with ethical principles governing responsible AI use.

The architecture of the model placed the mathematics teacher at the centre of the instructional process. GenAI functioned as a complementary learning assistant that provided personalised explanations, adaptive guidance, immediate feedback, and alternative solution strategies. Learners interacted with the system through carefully formulated prompts, evaluated the accuracy and logical consistency of generated responses, reflected on their reasoning, and received teacher feedback before accepting mathematical conclusions. Figure 2 presents the model's architecture.

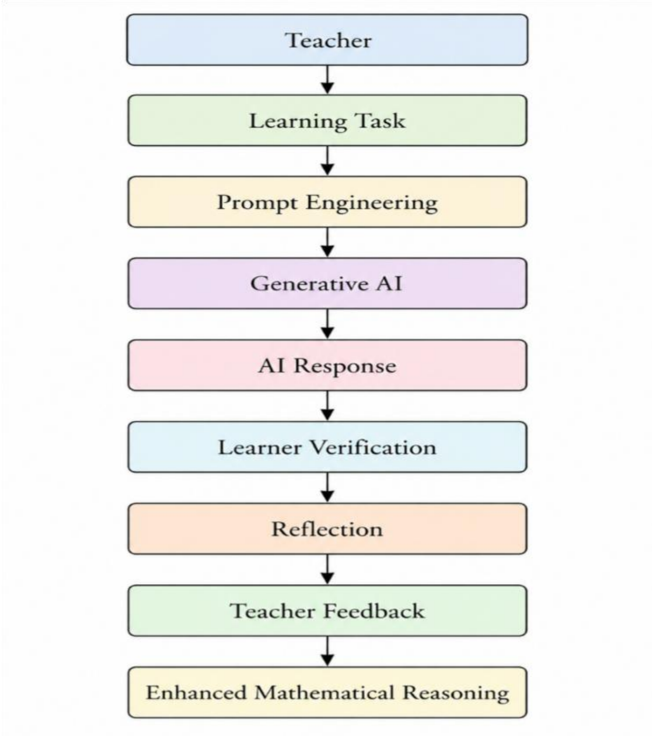

Figure 2: Architecture of the AI-Assisted Mathematics Learning Model (AIAMLM).

The instructional cycle began with the teacher introducing a curriculum-aligned mathematical concept or problem. Learners then formulated prompts and interacted with the GenAI system to obtain explanations, worked examples, hints, or alternative solution strategies. Rather than accepting the generated responses directly, learners checked their mathematical correctness, compared alternative approaches, and explained the reasoning underlying the proposed solutions. Questionable responses were discussed with peers and the teacher.

Prompt formulation was an important component of the model because the quality and relevance of GenAI responses depended partly on the clarity and specificity of the instructions provided. Learners were encouraged to use prompts that requested explanations, justification, guided hints, comparison of alternative methods, and identification of errors rather than prompts requesting final answers only. This approach was intended to maintain learners' cognitive engagement and promote higher-order reasoning (Kasneci et al., 2023; Biton et al., 2025).

During the verification stage, learners examined AI-generated solutions for correctness, logical consistency, completeness, and mathematical validity. They compared the generated explanations with their own reasoning, identified possible misconceptions, and consulted the teacher when necessary. This stage was intended to discourage uncritical reliance on GenAI and strengthen learners' analytical and evaluative skills (Kasneci et al., 2023; Walkington, 2025).

Learners subsequently engaged in structured reflection by evaluating their solution strategies, identifying misconceptions, comparing methods, and considering how the acquired knowledge could be transferred to related problems. Reflection supported metacognitive awareness and selfregulated learning by requiring learners to monitor and evaluate their mathematical thinking (Flavell, 1979; Zimmerman, 2002).

Teacher mediation remained essential throughout the cycle. Teachers introduced the learning tasks, guided prompt formulation, monitored learner–AI interactions, validated generated responses, facilitated mathematical discussion, and addressed misconceptions. The model consequently positioned GenAI as a complementary instructional resource rather than a replacement for the mathematics teacher.

3.9 Implementation of the Model

The AIAMLM was implemented in technology-supported secondary school mathematics classrooms in which learners had access to approved GenAI applications and internet-enabled devices. Lessons were aligned with the national secondary school mathematics curriculum and followed a learner-centred approach in which the teacher remained the principal instructional facilitator.

The lessons incorporated problem-solving tasks, collaborative discussion, reflective activities, and AI-assisted exploration. Learners used structured prompts to request explanations, justify procedures, verify generated responses, identify errors, and compare alternative solution strategies. They were discouraged from using GenAI merely to obtain final answers. Teachers monitored the interactions, facilitated classroom discussion, corrected misconceptions, and reinforced ethical and responsible use.

Implementation was evaluated through formative and summative assessment. Formative assessment included classroom reasoning activities, AI-assisted tasks, learner explanations, reflective exercises, and continuous teacher feedback. Summative assessment was based on the Mathematical Reasoning Test administered before and after the intervention. The assessment focused on learners' ability to analyse problems, construct logical arguments, justify strategies, communicate mathematical ideas, and apply mathematical concepts to unfamiliar situations.

4. Results

4.1 Participant Characteristics

A total of 130 participants were represented in the dataset, comprising 120 secondary school learners and 10 mathematics teachers. The learner sample included 60 participants in the experimental group and 60 in the control group. Among the learners, 58 (48.3%) were male and 62 (51.7%) were female. Their simulated mean age was M = 16.33 years (SD = 0.91, range = 15 −−18 years).

The teacher sample consisted of five male and five female mathematics teachers. Their mean age was M = 36.80 years (SD = 6.39, range = 28 −−49 years), and their teaching experience ranged from 2 to 18 years. Four teachers (40.0%) had between 6 and 10 years of teaching experience. Table 1 presents the demographic characteristics.

Table 1: Demographic characteristics of the participants

Characteristic	*n*	%
Learners (*N* = 120)		
Male	58	48.3
Female	62	51.7
Experimental group	60	50.0
Control group	60	50.0
Mean age, years	*M* = 16.33, *SD* = 0.91	
Age range, years	15 −−18	
Teachers (*N* = 10)		
Male	5	50.0
Female	5	50.0
Mean age, years	*M* = 36.80, *SD* = 6.39	
Age range, years	28 –49	
Teaching experience		
1–5 years	3	30.0
6–10 years	4	40.0
More than 10 years	3	30.0
Highest teaching qualification		
Bachelor's degree	5	50.0
Honours degree/postgraduate diploma	3	30.0
Master's degree	2	20.0
Other	0	0.0

N denotes the total sample size, *n* denotes frequency, *M* denotes the mean, and *SD* denotes the standard deviation.

4.2 Descriptive Results for Mathematical Reasoning

Table 2 presents the pre-test, post-test, and gain-score statistics. Before the intervention, the experimental group obtained a mean pre-test score of M = 50.30 (SD = 8.61), compared with M = 48.75 (SD = 6.52) for the control group. Following the eight-week intervention, the experimental group's mean score increased to M = 73.17 (SD = 7.75), whereas the control group obtained a post-test mean of M = 57.01 (SD = 6.64). The corresponding mean gains were M_{gain} = 22.87 (SD = 7.95) for the experimental group and M_{gain} = 8.26 (SD = 6.76) for the control group.

Table 2: Pre-test, post-test, and gain scores by instructional group

		Pre-test		Post-test		Gain score	
Group	n	M	SD	M	SD	M	SD
Experimental	60	50.30	8.61	73.17	7.75	22.87	7.95
Control 60	48.75	6.52	57.01	6.64	8.26	6.76	

Gain score = post-test score − pre-test score.

4.3 Baseline Comparison

An independent-samples *t*-test was used to compare the pre-test scores. The difference between the experimental group (M = 50.30, SD = 8.61) and the control group (M = 48.75, SD = 6.52) was not statistically significant, $t(118)$ = 1.11, p = .268, d = 0.20, 95% CI [−1.21,4.31]. The small effect size and nonsignificant difference indicated comparable baseline performance. Pretest performance was nevertheless retained as a covariate in the ANOVA.

4.4 Within-Group Changes in Mathematical Reasoning

Paired-samples *t*-tests were conducted to examine changes within each group. The experimental group improved from M = 50.30 (SD = 8.61) at pre-test to M = 73.17 (SD = 7.75) at post-test. This improvement was statistically significant, $t(59)$ = 22.30, p < .001, d = 2.88, representing a large standardised change.
The control group also improved from M = 48.75 (SD = 6.52) to M = 57.01 (SD = 6.64). The improvement was statistically significant, $t(59)$ = 9.46, p < .001, d = 1.22. Although both groups improved, the mean gain was substantially larger in the experimental group.

Table 3: Paired-samples *t*-tests for mathematical reasoning

Group	Mpre	SD_{pre}	Mpost	SDpost	t	df	p	d
Experimental	50.30	8.61	73.17	7.75	22.30	59	< .001	2.88
Control	48.75	6.52	57.01	6.64	9.46	59	< .001	1.22

Cohen's d represents the standardised within-group change.

4.5 Between-Group Post-Test Comparison

An independent-samples *t*-test showed that the experimental-group post-test score (M = 73.17, SD = 7.75) was significantly higher than the control-group score (M = 57.01, SD = 6.64), $t(118)$ = 12.27, p < .001, d = 2.24, 95% CI [13.55,18.77]. The standardised mean difference represented a large effect in favour of GenAI-assisted instruction.

4.6 ANOVA Results

ANOVA was conducted to evaluate the effect of instructional group on post-test mathematical reasoning after controlling for pre-test performance. The group-by-pre-test interaction was not statistically significant, $F(1,116) < 0.01$, p = .991, indicating that the homogeneity-ofregression-slopes assumption was satisfied.
After adjustment for pre-test performance, the effect of instructional group was statistically significant, $F(1,117)$ = 180.96, p < .001, partial η^2 = .607. The adjusted post-test mean was M_{adj} = 72.80 (SE = 0.81) for the experimental group and M_{adj} = 57.38 (SE = 0.81) for the control group.
The model explained 67.4% of the variance in post-test mathematical reasoning, R^2 = .674, adjusted R^2 = .668. The results therefore indicated a statistically significant and practically large instructional effect after adjustment for baseline performance.

Table 4: ANOVA of post-test mathematical reasoning scores

Source	SS	df	MS	F	p	η_p^2
Pre-test score	1582.72	1	1582.72	40.60	< .001	.258
Instructional group	7053.97	1	7053.97	180.96	< .001	.607
Error	4560.76	117	38.98			
Corrected total	13976.72	119				

The dependent variable was post-test mathematical reasoning, and the pretest score was the covariate. η_p^2 denotes partial eta squared.

Table 5: Adjusted post-test mathematical reasoning means

Group	n	M_{adj}	SE	95% CI	
				Lower	Upper
Experimental	60	72.80	0.81	71.20	74.40
Control	60	57.38	0.81	55.78	58.98

Adjusted means were estimated after controlling for pretest mathematical reasoning performance.

4.7 Hypothesis Testing

The principal null hypothesis stated that there would be no statistically significant difference in mathematical reasoning between learners receiving GenAI-assisted instruction and those receiving conventional instruction after controlling for pre-test performance.

The ANOVA showed a statistically significant instructional-group effect, $F(1,117) = 180.96$, $p < .001$, partial $\eta^2 = .607$. On the basis of the results, the null hypothesis was rejected. The illustrative conclusion was that GenAI-assisted instruction produced higher adjusted mathematical reasoning scores than conventional instruction.

4.8 Learner and Teacher Perceptions

The qualitative analysis generated five themes: personalised explanations and conceptual clarity, immediate feedback and strategy comparison, dependence and accuracy concerns, teacher mediation, and implementation challenges. The excerpts in Table 6 are synthetic statements made by participants.

Table 6: Themes concerning experiences of GenAI-assisted instruction

Theme	Source	Synthetic Illustrative Evidence
Personalised explanations and conceptual clarity	Learner response	“When I did not understand why the formula worked, the AI explained the same step in simpler words and gave me another example.” (L17)
Immediate feedback and strategy comparison	Learner response	“It showed me both substitution and elimination, so I could compare the methods and choose the one I understood better.” (L42)
Dependence and inaccurate responses	Teacher response	“Some learners initially wanted to copy the first answer, even when one of the calculation steps was incorrect.” (T03)
Teacher mediation and responsible use	Teacher response	“I still had to ask learners why the answer was correct and help them check every step before accepting it.” (Simulated T08)
Infrastructure and implementation challenges	Teacher response	“Unstable internet sometimes interrupted the activity, and several learners had to share one device.” (T06)

All excerpts are synthetic and were created solely to the reporting structure.

4.8.1 Personalised Explanations and Conceptual Clarity

In the analysis, learners valued explanations that responded directly to their areas of difficulty. One synthetic response stated:
“When I did not understand why the formula worked, the AI explained the same step in simpler words and gave me another example.” (learner L17)
Another learner response noted:
“I could ask the question again in a different way until the explanation became clear.” (learner L31)
These excerpts illustrate how personalised explanations could support conceptual clarity when learners actively questioned and evaluated the generated content.

4.8.2 Immediate Feedback and Alternative Strategies

The responses suggested that immediate feedback assisted learners in identifying errors and comparing solution procedures:
“It showed me both substitution and elimination, so I could compare the methods and choose the one I understood better.” (learner L42)
A synthetic observation entry similarly recorded:

"Three learners compared two AI-generated solution methods and explained to their group why the elimination method required fewer steps." (Observation O07, Session 5)
The convergence between learner responses and observation evidence illustrates how GenAI could promote strategy comparison rather than merely provide final answers.

4.8.3 Dependence and Accuracy Concerns

The qualitative evidence also illustrated concerns about learner dependence. One synthetic learner response stated:
"Sometimes I wanted to use the answer immediately instead of first trying the problem on my own." (learner L54)
A synthetic teacher response described an inaccurate output:
"The AI used the correct formula but made a sign error in the second step, and several learners did not notice it until we checked the solution together." (teacher T03)
These accounts demonstrate why verification and independent problem-solving should remain central to GenAI-assisted instruction.

4.8.4 Teacher Mediation

Teacher mediation emerged as a central theme. One synthetic teacher response stated:
"I still had to ask learners why the answer was correct and help them check every step before accepting it." (teacher T08)
A corresponding synthetic field note recorded:
"The teacher stopped the class after an incorrect AI-generated simplification and asked learners to locate and correct the error before continuing." (observation O11, Session 7)
These illustrative excerpts position the teacher as the person responsible for maintaining mathematical rigour and ensuring that GenAI supports rather than replaces learner reasoning.

4.8.5 Implementation Challenges

The analysis identified internet connectivity, device availability, teacher preparation, and lesson time as implementation challenges. One synthetic teacher response stated:
"Unstable internet sometimes interrupted the activity, and several learners had to share one device." (teacher T06)
This theme illustrates that the effectiveness of GenAI-assisted instruction may depend on reliable infrastructure, equitable device access, and adequate teacher preparation.

4.9 Classroom Observation Findings

For illustration, 16 experimental-group lessons were represented in the observation dataset. Active verification of AI-generated responses was recorded in 13 of the 16 lessons (81.3%), while explicit comparison of alternative solution strategies occurred in 11 lessons (68.8%). Teacher intervention to clarify or correct an AI-generated response was documented in nine lessons (56.3%).
A synthetic field note recorded:
"Learners questioned an AI-generated answer, recalculated the final step, and presented the corrected solution to the class." (observation O13, Session 8)

The observations also included negative cases. Uncritical acceptance of an AI-generated response occurred in four lessons (25.0%), while internet interruptions were documented in five lessons (31.3%). One synthetic field note stated:
"Two groups copied the generated solution without checking the substitution step; the teacher required them to solve the problem again without the AI response." (observation O04, Session 3)

The observation evidence therefore illustrates a balanced result: GenAI-assisted instruction may promote participation, comparison, and verification, but its educational value remains dependent on teacher mediation, reliable infrastructure, and learners' willingness to evaluate AI-generated content critically.

5. Discussion

5.1 Interpretation of the Quantitative Findings

The results indicated that structured Generative Artificial Intelligence (GenAI)-assisted instruction was associated with greater improvement in mathematical reasoning than conventional mathematics instruction. The experimental and control groups demonstrated comparable baseline performance, as the pre-test difference was not statistically significant, $t(118) = 1.11$, $p = .268$, $d = 0.20$. This baseline comparability reduced, although it did not eliminate, concern that the post-intervention difference resulted from initial differences between the groups.

Both groups demonstrated statistically significant pre-test–post-test improvement. The experimental group recorded a mean gain of 22.87 points, whereas the control group gained 8.26 points. Improvement in the control group may be attributed to continued curriculum exposure, regular classroom instruction, familiarity with the assessment format, and normal academic development during the intervention period. The larger improvement in the experimental group suggests that the structured use of GenAI provided additional learning support beyond that associated with conventional instruction.

The unadjusted post-test difference was statistically significant, $t(118) = 12.27$, $p < .001$, with a large standardised effect, $d = 2.24$. More importantly, ANOVA showed that the instructionalgroup effect remained statistically significant after adjustment for pre-test performance, $F(1,117) = 180.96$, $p < .001$, partial $\eta^2 = .607$. The adjusted post-test mean for the experimental group was 72.80, compared with 57.38 for the control group. Within the model, instructional group therefore accounted for a substantial proportion of the adjusted variation in mathematical reasoning.

The model explained 67.4% of the variance in post-test performance. Pre-test performance was also a significant covariate, $F(1,117) = 40.60$, $p < .001$, partial $\eta^2 = .258$, demonstrating that learners' prior mathematical knowledge remained relevant to their post-intervention performance. This result is educationally important because GenAI-assisted instruction should not be viewed as operating independently of learners' existing conceptual knowledge. Learners require sufficient foundational knowledge to formulate meaningful questions, interpret generated explanations, and detect errors in AI-generated solutions.

The effect sizes are unusually large and should be interpreted cautiously. Liu et al. (2026) , in a meta-analysis of 22 studies involving 5,232 participants, reported a moderate overall effect of GenAI on mathematics learning, $g = 0.534$, and a smaller effect among secondary school learners, $g = 0.313$. The between-group effect of $d = 2.24$ is therefore considerably larger than the effect normally observed across empirical studies. Such a difference could arise from intensive implementation, close teacher supervision, alignment between the intervention and assessment, short-term testing, or small-sample effects.

5.2 Interpretation of the Qualitative Findings

The qualitative analysis provided possible explanations for the quantitative pattern. Five themes were identified: personalised explanations and conceptual clarity, immediate feedback and strategy comparison, dependence and accuracy concerns, teacher mediation, and implementation challenges. These themes illustrate that the educational effect of GenAI may depend less on the mere availability of the technology than on how it is incorporated into mathematical activity.

The theme of personalised explanations and conceptual clarity suggested that learners valued the ability to request simpler explanations and additional examples. Unlike a fixed textbook explanation, a conversational GenAI system can reformulate a response when a learner does not understand the first presentation. This responsiveness may support conceptual understanding, particularly when learners are encouraged to connect generated explanations with their prior knowledge and classroom instruction (Kasneci et al., 2023; Walkington, 2025).

The second theme concerned immediate feedback and comparison of alternative strategies. The synthetic learner responses indicated that GenAI helped learners compare methods such as substitution and elimination rather than depend on a single prescribed procedure. From a mathematical reasoning perspective, the comparison of methods can help learners recognise structural relationships, evaluate procedural efficiency, and explain why different methods produce equivalent results. This interpretation is consistent with the argument that educational technology becomes most valuable when it supports reasoning and representation rather than merely performing calculations (Biton et al., 2025; Drijvers, 2023).

The classroom observations were consistent with these themes. Verification of AI-generated responses was documented in 13 of 16 lessons, while comparison of alternative strategies occurred in 11 lessons. These patterns illustrate how structured GenAI use could promote discussion, error analysis, and reflective thinking. However, uncritical acceptance of generated responses occurred in four lessons. Thus, the same technology that supports explanation and strategy comparison may also encourage passive answer copying when the instructional task does not require learners to justify and verify their conclusions.

Dependence and inaccurate responses consequently emerged as important counterthemes. The synthetic evidence included an example in which GenAI selected an appropriate formula but introduced a sign error during the calculation. This illustrates a central limitation of large language models: their responses may appear fluent and logically organised even when individual steps are mathematically invalid (Kasneci et al., 2023; Wardat et al., 2023). Learners with weak prior knowledge may be particularly vulnerable because they may lack the conceptual resources needed to detect plausible but incorrect explanations.

Teacher mediation therefore emerged as a necessary condition for effective implementation. In the observations, teachers questioned learners, required justification, corrected misconceptions, and prevented the direct acceptance of AI-generated answers. These practices positioned GenAI as a source of material for mathematical evaluation rather than an unquestioned authority. The finding supports a human–AI collaborative model in which the teacher retains responsibility for curriculum alignment, mathematical accuracy, classroom dialogue, and ethical use.

The implementation challenges, including unstable connectivity, device sharing, limited lesson time, and teacher preparation, further demonstrate that pedagogical potential does not guarantee equitable implementation. Where access to devices and internet connectivity is uneven, GenAI integration may widen rather than reduce educational inequalities. Infrastructure and professional development must therefore be considered integral components of implementation rather than secondary logistical concerns.

5.3 Comparison with Previous Studies

The pattern is broadly consistent with previous literature reporting that GenAI can support mathematical explanation, feedback, engagement, and problem-solving. Kasneci et al. (2023) argued that large language models can provide personalised assistance and immediate feedback, while also warning that inaccurate content and learner overdependence may undermine learning. Similarly, Wardat et al. (2023) found that ChatGPT could support mathematics teaching and learning through conversational explanations but required users to verify its mathematical accuracy.The cognitive benefits also correspond with the meta-analytic findings of Liu et al. (2026). That study reported a significant effect on cognitive mathematics outcomes, g = 0.596, and a larger effect on higher-order cognitive skills, g = 0.718. These findings provide support for the proposition that GenAI can contribute to reasoning-oriented outcomes when its use requires learners to analyse, compare, justify, and evaluate. Nevertheless, Liu et al. (2026) found that effects varied according to learning content, sample size, instructional mode, and the depth of GenAI integration. The present effect should therefore be interpreted as a possible outcome under structured implementation rather than a universal consequence of GenAI access.

The learner perceptions are comparable with the findings of Egara et al. (2025), who examined 125 secondary school learners in Nigeria. Learners valued ChatGPT for improving understanding, problem-solving, immediate assistance, and personalised learning, but also expressed concerns about its limitations and consistent use. The current analysis extends that perception-based pattern by illustrating how learner experiences could be examined alongside pre-test–post-test performance.

The emphasis on teacher mediation is consistent with the South African study by Motseki and Jojo (2025). Their study of 10 Grade 12 mathematics teachers found that ChatGPT could support geometry instruction and provide useful explanations, but it was less effective in correcting misconceptions. The accuracy of its responses depended on the cognitive demand of the problem, the input supplied, and prompt formulation. The findings similarly suggest that teachers must validate outputs and guide learners in distinguishing procedural plausibility from mathematical correctness.

The implementation themes also correspond with Namibian research. Elifas and Simuja (2024) found that AI-supported tools enhanced visualisation, paced learning, mathematical comprehension, and problem-solving in rural upper-primary classrooms. However, inadequate infrastructure, limited device access, insufficient teacher training, and a lack of culturally relevant tools restricted implementation. These challenges closely resemble the themes of connectivity, device sharing, and professional-development needs.

Likewise, Jatileni et al. (2024) found that Namibian teachers' intention to adopt AI was influenced by perceived relevance, attitudes, confidence, and perceptions of AI for social good. Stefanus et al. (2025) also reported positive teacher intentions but identified limited resources and low perceived behavioural control as barriers. Together, these studies reinforce the view that teacher acceptance is necessary but insufficient. Successful adoption also requires technological access, pedagogical competence, institutional support, and clear guidelines for responsible use.

The analysis differs from much of the existing Namibian literature by focusing on secondary school learners' mathematical reasoning rather than teacher adoption or general perceptions. It also illustrates the use of a comparison group and baseline-adjusted analysis. The study would extend Namibian evidence from questions of readiness and perceived usefulness towards the instructional effectiveness of GenAI-supported mathematics learning.

5.4 Relationship to the Theoretical Framework

The findings are compatible with the theoretical foundations of the study. Constructivist Learning Theory proposes that learners construct knowledge through active engagement, exploration, and reflection (Piaget, 1950; Bruner, 1960). In the intervention, learners did not merely receive AI-generated answers; they compared strategies, questioned explanations, and corrected errors. These activities illustrate active knowledge construction rather than passive information reception.

Social Constructivism similarly emphasises guided participation and scaffolding within the learner's Zone of Proximal Development (Vygotsky, 1978). GenAI provided supplementary explanations and hints, while teachers and peers supported interpretation, verification, and discussion. The results therefore position GenAI as one component within a broader social learning environment rather than as an independent tutor.

Cognitive Load Theory provides another possible explanation for the improvement. Stepby-step explanations and learner-appropriate examples may reduce extraneous cognitive load, allowing learners to focus on mathematical relationships (Sweller, 2020; Sweller et al., 2011). However, long, inaccurate, or unnecessarily complex responses may increase cognitive demand. Effective prompt formulation and teacher selection of suitable AI outputs are therefore important for maintaining an appropriate cognitive load.

The findings also align with Metacognitive Theory. Learners were required to monitor their understanding, evaluate generated procedures, identify errors, and reflect on alternative methods (Flavell, 1979; Zimmerman, 2002). Verification transformed the AI response into an object of analysis, thereby creating opportunities for metacognitive monitoring. Finally, the positive perceptions are compatible with the Technology Acceptance Model, although perceived usefulness and ease of use should not be interpreted as evidence of learning effectiveness without corresponding performance data (Davis, 1989; Venkatesh and Davis, 2000).

5.5 Implications for Mathematics Teaching

The findings suggest that GenAI should be incorporated as a structured reasoning scaffold rather than an unrestricted answer-generating tool. Mathematics teachers should design activities in which learners first attempt a problem independently, formulate explanatory prompts, compare generated strategies, verify each mathematical step, and justify the final conclusion. Such a sequence can preserve productive struggle while allowing learners to benefit from immediate and personalised support.

Teachers should also require learners to document how they used GenAI. For example, learners may submit their original attempt, the prompt used, the generated response, errors identified, corrections made, and a reflection on what they learned. This approach would make the reasoning process visible and reduce the likelihood of uncritical answer reproduction.

Assessment practices may also require revision. Tasks that ask only for final answers are increasingly vulnerable to automation. Greater emphasis should be placed on oral explanation, written justification, comparison of methods, error analysis, application to unfamiliar situations, and reflection on the validity of AI-generated solutions. These forms of assessment are more closely aligned with mathematical reasoning and provide stronger evidence of learners' independent understanding.

5.6 Implications for Curriculum Development

Curriculum developers should consider incorporating AI literacy into mathematics education. This should include prompt formulation, mathematical verification, recognition of AI limitations,

protection of personal information, acknowledgement of AI assistance, and responsible use. AI activities should be aligned with curriculum outcomes and should explicitly require reasoning, justification, and reflection.

Curriculum guidance should also distinguish appropriate assistance from cognitive substitution. Using GenAI to obtain a hint, examine an alternative representation, or critique an attempted solution may support learning. Using it to generate a complete response without learner engagement may undermine the intended outcome. Clear examples of acceptable and unacceptable classroom use would assist teachers in applying these distinctions consistently.

Implementation policies must also account for unequal access. Schools with limited connectivity or device availability may require shared-device activities, teacher-mediated demonstrations, downloadable resources, or offline alternatives. GenAI integration should not become a prerequisite for curriculum participation unless equitable access can be ensured.

5.7 Implications for Teacher Education

Teacher education and professional-development programmes should prepare mathematics teachers to evaluate GenAI outputs both pedagogically and mathematically. Training should address the design of reasoning-oriented prompts, identification of hallucinations and computational errors, differentiation of instruction, management of learner dependence, data privacy, academic integrity, and assessment redesign.

Professional development should move beyond demonstrating how to operate a GenAI application. Teachers require opportunities to design AI-assisted lessons, evaluate generated solutions, observe modelled classroom practice, and reflect collaboratively on implementation challenges. The Technological Pedagogical Content Knowledge framework may provide a useful basis for integrating mathematical content knowledge, pedagogy, and technological competence.

Teacher education should also reinforce that GenAI does not reduce the importance of the teacher. The simulated findings instead suggest that effective GenAI use increases the need for skilled teacher mediation. Teachers must determine when AI support is appropriate, ensure the accuracy of mathematical content, maintain learner engagement, and protect opportunities for independent reasoning.

5.8 Policy Implications

Educational policymakers should develop context-sensitive guidelines governing the use of GenAI in schools. These guidelines should address data protection, age-appropriate access, academic integrity, transparency, teacher responsibility, acceptable assessment use, and procedures for evaluating AI-generated content. Policies should permit pedagogically valuable experimentation while establishing safeguards against misuse.

Investment in connectivity, devices, technical support, and teacher professional development is also necessary. Without these resources, the benefits of GenAI-assisted learning may remain concentrated in better-resourced schools. Policy development should therefore treat equity, infrastructure, and teacher capacity as central conditions for responsible AI integration.

Overall, the simulated findings illustrate how structured GenAI-assisted instruction could strengthen mathematical reasoning when learners actively question, compare, verify, and reflect on generated responses. They also show that technology alone is insufficient: the anticipated benefit depends on teacher mediation, sound task design, reliable infrastructure, ethical guidance, and learners' continued engagement in independent mathematical thinking.

6. Limitations

This study has several limitations. The use of intact classes within a quasi-experimental nonequivalent control group design limited individual randomisation and, despite adjustment for pre-test performance through ANOVA, unmeasured group differences may have influenced the results. The purposive selection of technologically prepared schools, teachers, and classes also limits the generalisability of the findings to less-resourced educational settings. Furthermore, the eight-week intervention measured short-term effects and could not establish the long-term sustainability of improvements in mathematical reasoning. Variations in internet connectivity, device access, digital literacy, teacher competence, and GenAI model performance may have affected implementation consistency and reproducibility. Finally, because GenAI systems may generate inaccurate or logically inconsistent mathematical responses, the effectiveness of the intervention depended partly on learners' verification skills and teachers' capacity to identify and correct AI-generated errors. These limitations require cautious interpretation of the findings.

7. Future Research

Future studies should compare GenAI platforms, teacher-guidance levels, prompt strategies, and mathematical domains, with emphasis on higher-order outcomes such as justification, proof, error analysis, and mathematical communication. Research should also examine how prior achievement, AI literacy, self-regulation, teacher competence, and technological access influence the effectiveness of GenAI-assisted instruction. Further investigation is needed into the accuracy and reproducibility of AI-generated mathematical explanations, as well as algorithmic bias, learner dependence, equitable access, and contextually appropriate policies for responsible GenAI integration in Namibian schools.

8. Conclusion

This study developed an AI-Assisted Mathematics Learning Model that positions GenAI as a complementary resource for providing explanations, feedback, and alternative solution strategies while preserving the teacher's central pedagogical role. The analysis produced a significant adjusted instructional-group effect, $F(1, 117) = 180.96$, $p < .001$, partial $\eta^2 = .607$, and illustrated potential benefits related to personalised support, strategy comparison, and reflective problemsolving. It also highlighted concerns regarding inaccurate responses, learner dependence, technological access, and the need for teacher mediation. Conceptually, the model suggests that GenAI can support mathematical reasoning when learners formulate purposeful prompts, verify generated solutions, justify conclusions, and reflect on their reasoning. The framework therefore provides a foundation for future empirical evaluation and responsible GenAI integration in Namibian secondary school mathematics education.

Acknowledgements

The authors sincerely thank the participating secondary schools, mathematics teachers, and learners for their valuable time, cooperation, and contributions to this study. Their willingness to participate and share their experiences made this research possible. The authors also acknowledge the constructive support provided by colleagues and experts in mathematics education whose insights contributed to the development of the AI-Assisted Mathematics Learning Model presented in this study. Finally, the authors express their appreciation to the International University of Management (IUM), Namibia, for providing an academic environment that supports research and innovation in mathematics education and educational technology.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest. The authors have no financial, personal, academic, or professional relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available because they contain information that could compromise the privacy and confidentiality of the research participants.

References

Anthropic. (2024). Claude 3 model card. https://www.anthropic.com/news/claude-3-family

Biton, Y., Segal, R., and Alush, K. (2025). How utilizing generative AI when addressing pedagogical and mathematical events contributes to mathematics teacher educators' TPACK. International Journal of Education in Mathematics, Science and Technology, 13(4), 895–913.

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bruner, J. S. (1960). The process of education. Harvard University Press.

Campbell, D. T., and Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Houghton Mifflin.

Chen, Y., Wang, H., and Lin, J. (2025). Generative artificial intelligence in mathematics education: Opportunities, challenges, and future directions. Computers and Education: Artificial Intelligence, 7.

Creswell, J. W., and Creswell, J. D. (2023). Research design: Qualitative, quantitative, and mixed methods approaches (6th ed.). SAGE Publications.

Creswell, J. W., and Plano Clark, V. L. (2024). Designing and conducting mixed methods research (4th ed.). SAGE Publications.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

Denny, P., et al. (2024). Generative AI for education (GAIED): Advances, opportunities, and challenges. arXiv. https://arxiv.org/abs/2402.01580

Drijvers, P. (2023). Digital technology in mathematics education: Why it works (or does not). Educational Studies in Mathematics, 113, 1–24.

Egara, F. O., Mosimege, M., and Mosia, M. (2025). Secondary school students' perceptions of their usage of artificial intelligence-based ChatGPT in mathematics learning. Journal of Education, 98. https://doi.org/10.17159/2520-9868/i98a07

Elifas, T. N., and Simuja, C. (2024). The pedagogical insights of mathematics teachers integrating artificial intelligence in rural upper primary schools of Namibia. Interdisciplinary Journal of Education Research, 6, 1–14. https://doi.org/10.38140/ijer-2024.vol6.37

Field, A. (2024). Discovering statistics using IBM SPSS Statistics (6th ed.). SAGE Publications.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. American Psychologist, 34(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Giannakos, M. N., et al. (2025). The promise and challenges of generative artificial intelligence in education. Behaviour & Information Technology, 44(3), 411–430.

Google DeepMind. (2024). Gemini: A family of highly capable multimodal models. https://arxiv.org/abs/2312.11805

Hattie, J. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. Routledge.

Holmes, W., Bialik, M., and Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.

Hoyles, C., and Lagrange, J.-B. (Eds.). (2010). Mathematics education and technology: Rethinking the terrain. Springer.

Ifenthaler, D., and Schumacher, C. (2022). Learning analytics, artificial intelligence, and digital transformation in education. Computers in Human Behavior, 128, Article 107093.

Jatileni, C. N., Sanusi, I. T., Olaleye, S. A., Ayanwale, M. A., Agbo, F. J., and Oyelere, P. B. (2024). Artificial intelligence in compulsory level of education: Perspectives from Namibian in-service teachers. Education and Information Technologies, 29(10), 12569–12596. https://doi.org/10.1007/s10639-023-12341-z

Johnson, R. B., and Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. Educational Researcher, 33(7), 14–26. https://doi.org/10.3102/0013189X033007014

Kasneci, E., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kilpatrick, J., Swafford, J., and Findell, B. (Eds.). (2001). Adding it up: Helping children learn mathematics. National Academies Press. https://doi.org/10.17226/9822

Kirschner, P. A., Sweller, J., and Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educational Psychologist, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Liu, B., Zhang, W., and Wang, F. (2026). Can generative artificial intelligence effectively enhance students' mathematics learning outcomes? A meta-analysis of empirical studies from 2023 to 2025. Education Sciences, 16(1), Article 140. https://doi.org/10.3390/educsci16010140

Mason, J., Burton, L., and Stacey, K. (2010). Thinking mathematically (2nd ed.). Pearson.

Miao, F., and Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Microsoft. (2024). Microsoft Copilot for education: Transforming teaching and learning. Microsoft Education.

Motseki, P., and Jojo, Z. (2025). Teachers' perceptions on the use of ChatGPT in teaching Grade 12 mathematics. Journal of Education, 99. https://doi.org/10.17159/2520-9868/i99a07

Namibia Ministry of Education, Arts and Culture. (2023). National curriculum for basic education.

Nanda, S. B., and Pradhan, D. K. (2025). Artificial intelligence in mathematics education: A systematic review of global trends and emerging themes. Journal of Mathematics Science and Technology Education, 2(2), 153–175.

National Council of Teachers of Mathematics. (2014). Principles to actions: Ensuring mathematical success for all.

Nunnally, J. C., and Bernstein, I. H. (1994). Psychometric theory (3rd ed.). McGraw–Hill.

OpenAI. (2023). GPT-4 technical report. arXiv. https://arxiv.org/abs/2303.08774

Organisation for Economic Co-operation and Development. (2023a). PISA 2022 results: The state of learning and equity in education (Vol. 1). OECD Publishing.

Organisation for Economic Co-operation and Development. (2023b). PISA 2022 mathematics framework. OECD Publishing.

Piaget, J. (1950). The psychology of intelligence. Routledge.

Polya, G. (1957). How to solve it (2nd ed.). Princeton University Press.

Russell, S. J., and Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.

Schoenfeld, A. H. (1985). Mathematical problem solving. Academic Press.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. Mathematics Teaching, 77, 20–26.

Stefanus, L., Nhinda, G. T., and Shaanika, I. N. (2025). Teachers' AI adoption in Namibia: A theory of planned behaviour analysis. Higher Education Research, 10(3), 77–87. https://doi.org/10.11648/j.her.20251003.12

Sweller, J. (2020). Cognitive load theory and educational technology. Educational Technology Research and Development, 68, 1–16.

Sweller, J., Ayres, P., and Kalyuga, S. (2011). Cognitive load theory. Springer.

UNESCO. (2021a). Reimagining our futures together: A new social contract for education.

UNESCO. (2021b). Recommendation on the ethics of artificial intelligence.

Venkatesh, V., and Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. Management Science, 46(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425–478. https://doi.org/10.2307/30036540

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Walkington, C. (2025). The implications of generative artificial intelligence for mathematics education. School Science and Mathematics, 125(2), 95–112.

Wardat, Y., Tashtoush, M. A., AlAli, R., and Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. Eurasia Journal of Mathematics, Science and Technology Education, 19(7), Article em2286.

https://doi.org/10.29333/ejmste/13272
World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. JAMA, 310(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053
Zawacki-Richter, A., Marín, V. I., Bond, M., and Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? International Journal of Educational Technology in Higher Education, 16, Article 39. https://doi.org/10.1186/s41239-019-0171-0
Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory Into Practice, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2