International Journal of Research Studies in Education
2026 Volume 15 Number 23, 157-173



Artificial Intelligence chatbot use as a learning aid and its relationship with learning comprehension among BSED science students

Moya, Ar-Jay M. ✉
Central Bicol State University of Agriculture (Pasacao Campus), Philippines (arjay.moya@cbsua.edu.ph)
Panaligan, Bernadeth R.
Central Bicol State University of Agriculture (Pasacao Campus), Philippines (bernadeth.panaligan@cbsua.edu.ph)
Calugay, Aires N.
Central Bicol State University of Agriculture (Pasacao Campus), Philippines (aires.calugay@cbsua.edu.ph)
Bello, Ma. Suseth S.
Central Bicol State University of Agriculture (Pasacao Campus), Philippines (masuseth.bello@cbsua.edu.ph)
Bonador, Jerico M.
Central Bicol State University of Agriculture (Pasacao Campus), Philippines (jerico.bonador@cbsua.edu.ph)
Aureada, Claire C.
Central Bicol State University of Agriculture (Pasacao Campus), Philippines (claire.aureada@cbsua.edu.ph)

ISSN: 2243-7703
Online ISSN: 2243-7711

OPEN ACCESS

Received: 3 November 2026 **Revised**: 13 November 2026 **Accepted**: 15 November 2026
Available Online: 16 November 2026 **DOI**: 10.5861/ijrse.2026.26460

Abstract

This study examined the relationship between artificial intelligence (AI) chatbot use as a learning aid and the perceived learning comprehension of first-year to third-year Bachelor of Secondary Education (BSED) Science students at Central Bicol State University of Agriculture–Pasacao Campus during Academic Year 2025–2026. It assessed AI chatbot use (concept explanation, problem-solving support, and review and reinforcement), perceived learning comprehension (understanding, retention, and application of scientific concepts), and their statistical relationship to inform instructional recommendations. Using a quantitative descriptive-correlational design with open-ended qualitative inquiry, data were gathered from 158 students selected via proportionate stratified random sampling from 260 students (N=260) through a validated 30-item Likert questionnaire on Google Forms (Cronbach's α=.951). Data were analyzed using weighted mean, ranking, Pearson's r, t-test, coefficient of determination (r^2), and semantic thematic analysis. Overall AI chatbot use was High (M=4.04), led by concept explanation (M=4.22, Very High), problem-solving support (M=4.04, High), and review and reinforcement (M=3.85, High). Perceived learning comprehension was also High (M=3.83) across understanding (M=3.84), application (M=3.83), and retention (M=3.82). Correlation analysis revealed a significant, moderate positive relationship between AI chatbot use and perceived learning comprehension, r(156)=.470, t=6.651, p<.001, r^2=.221, explaining 22.1% of comprehension variance. Thematic analysis corroborated these findings while

© The Author(s) / Attribution-NonCommercial-NoDerivs CC BY-NC-ND

highlighting output accuracy and overdependence concerns, concluding that AI chatbots serve as effective scaffolds when paired with source verification, independent reasoning, and faculty guidance.

Keywords: artificial intelligence chatbots, learning aid, learning comprehension, preservice science teachers, science education, higher education

Artificial Intelligence chatbot use as a learning aid and its relationship with learning comprehension among BSED science students

1. Introduction

Artificial intelligence–driven chatbots have transitioned rapidly from experimental curiosities to everyday companions in academic life. In higher education, particularly in science teacher education, their presence is increasingly evident when learners require immediate academic support—when making sense of abstract scientific principles, navigating multi-step problem-solving, or reinforcing concepts outside limited classroom hours. Generative AI systems capable of answering disciplinary questions, paraphrasing technical terminology, and simulating tutorial dialogue now operate alongside textbooks, laboratory manuals, and learning management systems. This technological shift raises a central pedagogical question for higher education: how do AI chatbots function as supplementary learning aids, and how does their utilization relate to students' learning comprehension rather than merely to task convenience or completion speed?

At the theoretical level, instructional technologies are expected to scaffold conceptual understanding without displacing students' cognitive effort. From a social constructivist perspective, learning tools should assist learners in actively constructing knowledge rather than passively receiving ready-made answers (Vygotsky, 1978). Vygotsky's concept of the Zone of Proximal Development (ZPD) posits that learners benefit from structured scaffolds that enable them to accomplish cognitive tasks slightly beyond their current independent capability. In science education, AI chatbots can provide immediate dialogic support to help students clarify misconceptions, verify procedural steps, and consolidate scientific concepts. However, persistent concerns remain regarding the factual accuracy of chatbot-generated explanations and the extent to which students critically evaluate automated outputs before accepting them. Consequently, the empirical relationship between AI chatbot use and science learning comprehension warrants systematic investigation.

Recent international literature documents the widespread adoption of generative AI chatbots across higher education disciplines. Empirical syntheses indicate that conversational agents such as ChatGPT are frequently utilized for concept clarification, academic writing support, programming, and structured problem-solving. A comprehensive systematic review and meta-analysis of experimental studies demonstrated that ChatGPT-supported interventions generally yield positive effects on students' academic performance, affective engagement, and higher-order thinking skills (Deng et al., 2025). Similarly, in controlled experimental settings, university students who utilized ChatGPT during complex problem-solving demonstrated enhanced solution quality, greater elaboration, and higher perceived self-efficacy, although researchers cautioned that perceived ease of AI use can sometimes lead learners to overestimate their independent mastery (Urban et al., 2024). Within STEM contexts, related studies further indicate that AI-powered tutoring systems can function as accessible virtual guides that deliver step-by-step explanations tailored to individual learner pacing (Gkintoni et al., 2025).

Beyond ChatGPT, conversational AI platforms such as Google Gemini and Microsoft Copilot are increasingly embedded within everyday academic workflows. Because these systems are integrated directly into web browsers, search engines, and productivity suites, they offer low-barrier access for information retrieval, concept explanation, and study organization. Multidisciplinary analyses emphasize that the ubiquity and conversational fluency of large language models have accelerated their adoption as informal supplementary learning aids among university students (Dwivedi et al., 2023; Kasneci et al., 2023). Consequently, contemporary learners rarely restrict themselves to a single platform; instead, they engage fluidly with multiple AI chatbot ecosystems depending on task demands and accessibility.

Within science and mathematics education specifically, empirical research on chatbot-assisted learning has yielded promising yet nuanced findings. In a quasi-experimental study of fifth-grade science learners, chatbot-

supported instruction did not produce statistically significant immediate gains in standardized achievement test scores, yet students reported highly positive learning experiences and valued the chatbot for lesson review and conceptual reinforcement (Deveci Topal et al., 2021). In the Philippine secondary education context, significant post-test improvements and stronger self-regulated learning behaviors were observed among Grade 10 students who utilized an interactive science chatbot as a supplementary review tool during distance learning (Prondoza & Panoy, 2022). Similarly, in Philippine teacher education, Alvarez et al. (2024) demonstrated that integrating AI-powered tutors (Flexi 2.0 and MathGPT) among preservice educators significantly enhanced personalized problem-solving and higher-order thinking, provided that instructors enforced strict ethical monitoring and required learners to critically evaluate AI-generated solutions. Together, these findings suggest that conversational AI tools can meaningfully support conceptual review and learner engagement when accompanied by structured pedagogical oversight.

Nevertheless, empirical findings regarding the academic impact of AI chatbots remain mixed. Recent quasi-experimental evidence indicates that while university students overwhelmingly perceive AI chatbots as convenient and helpful study aids, unguided usage does not automatically translate into superior learning outcomes across all assessment dimensions (Eteng-Uket & Ezeoguine, 2025). Such divergences suggest that the educational value of AI chatbots depends critically on how learners engage with them—whether as cognitive scaffolds for understanding, procedural guides for problem-solving, or active tools for retrieval practice—rather than on their mere availability. Examining the specific functional dimensions of AI chatbot use and their relationship with students' perceived learning comprehension is therefore essential.

Despite the rapid expansion of global AI-in-education scholarship, notable contextual gaps persist. Much of the extant empirical literature examines students in well-resourced, urban, or Western universities. Comparatively few investigations have focused on preservice science teachers—specifically Bachelor of Secondary Education (BSED) major in Science students—enrolled in regional state universities and colleges in developing countries. In the Philippine higher education landscape, technological integration varies across regional campuses due to bandwidth constraints, device availability, and evolving institutional policies. Recent regional evidence from the Bicol region indicates that university students actively employ generative AI as a cognitive scaffold under academic time pressure, yet often experience an ethical paradox regarding proper disclosure and independent mastery (Gile & Fortes, 2026). Moreover, limited local empirical research has disaggregated how specific functional modes of AI chatbot use relate to distinct cognitive dimensions of science learning comprehension.

This gap is particularly salient at Central Bicol State University of Agriculture–Pasacao Campus, where first-year to third-year BSED Science students navigate rigorous coursework spanning biology, chemistry, physics, and earth science alongside pedagogical training. When confronted with dense scientific theories, quantitative problem sets, and examination preparation under time and resource constraints, students increasingly turn to AI chatbots as informal study aids. While some students report that chatbot explanations sharpen their grasp of difficult topics, informal classroom observations also raise concerns that passive reliance on AI-generated outputs may foster superficial answer copying rather than genuine conceptual mastery. Anchored in social constructivist scaffolding and cognitive load regulation, this study advances the central hypothesis that the extent of AI chatbot use as a learning aid is significantly associated with the degree of perceived learning comprehension among BSED Science students—a proposition tested directly by the study's third research question. Addressing this empirical gap carries direct pedagogical and institutional significance: it equips preservice science teachers with empirical self-awareness regarding their study habits, enables science faculty to design targeted formative assessments and retrieval practice, and provides academic administrators with localized baseline evidence for crafting institutional AI literacy policies.

Theoretical and Conceptual Framework - This study is anchored in complementary theoretical perspectives spanning technology adoption, cognitive psychology, and instructional design to explain how AI chatbot use as a learning aid relates to students' perceived learning comprehension (see Figure 1). At the adoption level, Davis's (1986) Technology Acceptance Model (TAM) posits that an individual's utilization of an information system is

governed by perceived usefulness (the belief that using the system enhances task performance) and perceived ease of use (the belief that using the system is free of undue effort). Within the present investigation, TAM explains why BSED Science students turn to accessible conversational AI platforms—specifically ChatGPT, Google Gemini, and Microsoft Copilot—for concept explanation, problem-solving support, and lesson review when navigating demanding science coursework.

At the cognitive and interactional level, Vygotsky's (1978) Social Constructivism and Zone of Proximal Development (ZPD) frame learning as a dialogically mediated process in which structured scaffolds enable learners to accomplish tasks slightly beyond their independent capability. Constructivist and situative scholarship further emphasizes that dialogic elaboration, guided questioning, and active knowledge construction allow learners to refine conceptual schemas that might remain inaccessible through solitary textbook reading alone (Greeno, 1997; Van Meter & Stevens, 2000). Complementing this perspective, Cognitive Load Theory (CLT) posits that human working memory is strictly limited when processing novel, high-interactivity material, whereas long-term memory stores organized cognitive schemas (Sweller, 1988; Sweller et al., 2011). In science education, conversational AI tools can regulate cognitive load by segmenting complex analytical tasks into manageable steps and reducing extraneous search effort, thereby freeing working memory capacity for germane schema construction across understanding, retention, and application (Gkintoni et al., 2025).

Finally, to translate empirical findings into actionable classroom practice, the study draws upon the Technological Pedagogical Content Knowledge (TPACK) framework (Koehler & Mishra, 2009; Mishra & Koehler, 2006), the ADDIE instructional design model (Branch, 2009; Molenda, 2003), and Uden's (2024) Activity Theory–based ecosystem for Artificial Intelligence in Education (AIED). While TPACK and ADDIE align digital tools with discipline-specific science pedagogy and systematic instructional design, Uden (2024) emphasizes that sustainable AI integration requires collaborative socio-cultural agreements among learners, educators, and institutions so that AI mediates rather than replaces human cognitive activity. Conceptually, as illustrated in Figure 1, the investigation examines how the independent variable—the extent of AI chatbot use across concept explanation, problem-solving support, and review and reinforcement—relates to the dependent variable—students' perceived learning comprehension across understanding of scientific concepts, retention of information, and application of scientific concepts—and triangulates these quantitative associations with qualitative student reflections to inform pedagogical governance.

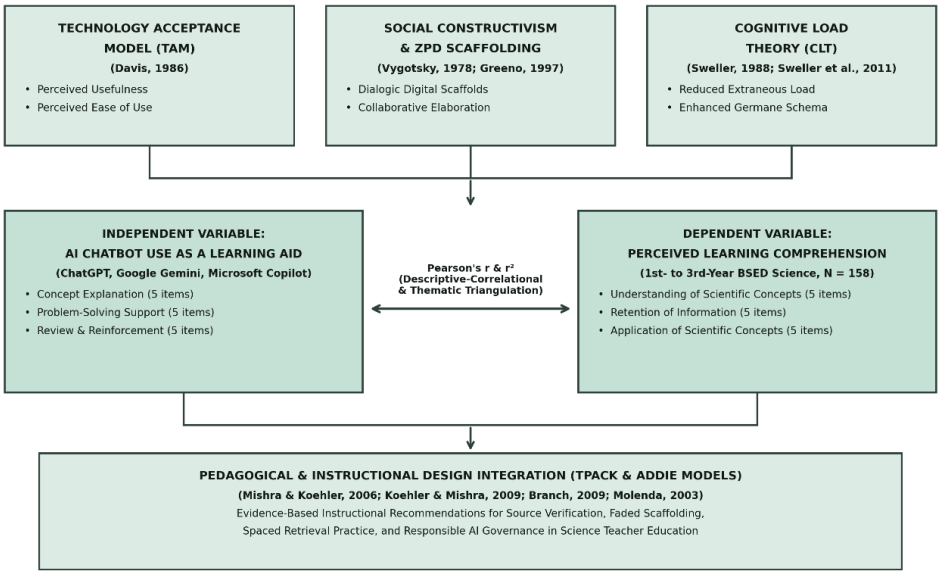


Figure 1. Theoretical and conceptual paradigm of the study.

Research Questions and Hypothesis - This study examined the relationship between artificial intelligence

(AI) chatbot use as a learning aid and the perceived learning comprehension of first-year to third-year BSED Science students at Central Bicol State University of Agriculture–Pasacao Campus during Academic Year 2025–2026. Specifically, it sought to answer the following research questions:

- To what extent do first-year to third-year BSED Science students use AI chatbots as a learning aid in terms of: (a) concept explanation, (b) problem-solving support, and (c) review and reinforcement of scientific concepts?
- To what degree do these BSED Science students perceive their learning comprehension in terms of: (a) understanding of scientific concepts, (b) retention of information, and (c) application of scientific concepts?
- Is there a statistically significant relationship between the extent of AI chatbot use as a learning aid and the degree of perceived learning comprehension among BSED Science students?
- What practical instructional recommendations and educational implications can be proposed based on the findings to guide the effective and responsible use of AI chatbots in science education?

Directly linked to the third research question (Research Question 3), the study tested the following hypotheses at the two-tailed $\alpha = .05$ level of significance: the null hypothesis (H_0) posited that there is no statistically significant relationship between the extent of AI chatbot use as a learning aid and the degree of perceived learning comprehension among first-year to third-year BSED Science students, whereas the alternative hypothesis (H_a) posited that a statistically significant positive relationship exists between the two constructs.

2. Methodology

Research Design - This study employed a non-experimental quantitative descriptive-correlational research design supplemented by open-ended qualitative inquiry (Creswell & Creswell, 2018). The rationale for selecting a descriptive-correlational design is grounded in three methodological considerations. First, the descriptive component was necessary to establish an accurate, unmanipulated baseline profile of two naturally occurring educational phenomena across three undergraduate cohorts: (a) the frequency with which preservice science teachers engage commercial AI chatbots across specific study functions, and (b) their self-assessed degree of comprehension across distinct cognitive domains. Because students already integrate conversational AI fluidly into their daily off-campus study routines, artificial laboratory restrictions would distort authentic usage patterns. Second, the correlational component was methodologically appropriate because the primary inferential objective (Research Question 3) was to quantify the statistical direction, bivariate magnitude (r), and proportion of shared variance (r^2) between AI chatbot utilization and perceived learning comprehension across a representative cross-section of learners (Field, 2018), where randomly assigning students to an AI-deprivation control group across an entire academic year was neither pedagogically equitable nor logistically feasible. Third, embedding three open-ended qualitative prompts within the survey architecture enabled methodological triangulation, allowing the researchers to explain *why* specific utilization patterns and comprehension gaps emerged through students' lived verification practices and accuracy concerns.

Population and Sampling Design - The target population comprised all 260 first-year, second-year, and third-year Bachelor of Secondary Education (BSED) major in Science students officially enrolled at Central Bicol State University of Agriculture–Pasacao Campus, Camarines Sur, Philippines, during Academic Year 2025–2026. Based on official enrollment records, the population was distributed across three year-level strata: 87 first-year students, 91 second-year students, and 82 third-year students. Fourth-year BSED Science students were excluded from the sampling frame because they were deployed off-campus for full-time student teaching internships during the data collection period, and non-Science education majors were likewise excluded.

Using finite-population sample size determination (Slovin's formula at a 5% margin of error), the required

sample size from the population of $N = 260$ was established at $n = 158$ respondents, which also provides greater than .99 statistical power to detect a moderate bivariate correlation of $r = .30$ at two-tailed $\alpha = .05$. To ensure equitable representation across academic cohorts, proportionate stratified random sampling was employed, yielding 53 first-year respondents (33.54%), 55 second-year respondents (34.81%), and 50 third-year respondents (31.65%). Individual participants within each stratum were selected via simple random sampling. As shown in Table 1, the actual number of completed responses matched the required stratum quotas exactly.

Table 1

Population and Sample Distribution of the Respondents by Year Level

Year Level	Population (N)	Required Sample (n)	Actual Respondents (n)	Percentage (%)
First Year	87	53	53	33.54
Second Year	91	55	55	34.81
Third Year	82	50	50	31.65
Total	260	158	158	100.00

Note. The required sample size (n = 158) was determined from the finite population (N = 260) at a 5% margin of error and allocated via proportionate stratified random sampling.

Research Instrument - Primary data were collected using a validated, four-part researcher-made questionnaire administered electronically via Google Forms. Part I recorded the respondents' year-level cohort to verify proportional stratum representation. Part II measured the independent variable—the extent of AI chatbot use as a learning aid (encompassing commonly accessed generative conversational platforms such as ChatGPT, Google Gemini, and Microsoft Copilot)—across 15 behavioral statements divided equally into three five-item functional dimensions: (a) *concept explanation*, defined operationally as the use of chatbots to clarify abstract scientific terminology, unpack complex theories, and provide illustrative examples; (b) *problem-solving support*, defined as procedural scaffolding to analyze scientific questions, trace step-by-step solution pathways, and verify reasoning when stuck; and (c) *review and reinforcement of scientific concepts*, defined as summarizing lessons, generating practice questions, and consolidating previously studied material.

Part III assessed the dependent variable—perceived learning comprehension—across 15 self-report statements divided equally into three five-item cognitive dimensions: (a) *understanding of scientific concepts*, defined operationally as the perceived capacity to comprehend scientific explanations, identify core principles, recognize relationships among ideas, and articulate concepts in one's own words; (b) *retention of information*, defined as the perceived ability to encode, store, and retrieve learned terminology, processes, and concepts during examinations and over time; and (c) *application of scientific concepts*, defined as the perceived ability to transfer learned principles to novel situations, real-life problems, and independent analytical tasks. Part IV comprised three open-ended prompts inviting respondents to describe how AI chatbots supported their science learning, what inaccuracies or challenges they encountered, and how AI chatbots should be used responsibly.

The 30 quantitative items in Parts II and III were rated on a five-point Likert scale ranging from 1 (*Never* for Part II utilization frequency; *Strongly Disagree* for Part III perceived comprehension) to 5 (*Always* / *Strongly Agree*). Content and face validity were established prior to administration through expert review conducted by faculty specialists in science education and educational research. A pilot test was subsequently conducted with $n = 30$ non-sample education students to evaluate internal consistency reliability. The 30-item instrument yielded an overall Cronbach's alpha coefficient of $\alpha = .951$, exceeding the standard .70 psychometric threshold (Field, 2018) and confirming high internal consistency.

Data Gathering Procedure - Data collection followed a standardized institutional protocol. First, the researchers submitted a formal letter of request to the Campus Administrator of CBSUA–Pasacao and the relevant College of Education faculty instructors to secure authorization to survey the first-year to third-year BSED Science cohorts. Upon receiving administrative approval, the researchers coordinated with course instructors and class presidents to distribute the survey asynchronously via official class communication channels without disrupting formal classroom or laboratory instruction. Prospective respondents were presented with an electronic informed consent briefing on the opening screen of the Google Form detailing the study's academic purpose, voluntary

participation principles, right to withdraw without penalty, and confidentiality safeguards. Following data collection, the 158 completed responses were audited for completeness, coded into a statistical spreadsheet, and subjected to quantitative and qualitative analyses.

Data Analysis - Quantitative data were analyzed at the two-tailed $\alpha = .05$ level of significance following behavioral science conventions (Field, 2018). Weighted mean (*M*) and ranking were computed to describe the extent of AI chatbot use (Part II) and the level of perceived learning comprehension (Part III). Item and dimensional weighted means were interpreted using a uniform five-point interval continuum: 4.21–5.00 (Scale 5, *Always* / *Strongly Agree*; Descriptive Interpretation: *Very High*); 3.41–4.20 (Scale 4, *Oftentimes* / *Agree*; *High*); 2.61–3.40 (Scale 3, *Sometimes* / *Moderately Agree*; *Moderate*); 1.81–2.60 (Scale 2, *Rarely* / *Disagree*; *Low*); and 1.00–1.80 (Scale 1, *Never* / *Strongly Disagree*; *Very Low*). To test the third research question, Pearson's product–moment correlation coefficient (*r*) was computed between the composite mean score of AI chatbot use and the composite mean score of perceived learning comprehension, accompanied by its 95% confidence interval, a two-tailed *t*-test for correlation significance ($df = N - 2 = 156$), and the coefficient of determination (r^2) to quantify the proportion of variance explained. Qualitative responses to the three open-ended prompts in Part IV were analyzed using semantic thematic analysis (Braun & Clarke, 2006) through iterative reading, initial semantic coding, and collation into overarching themes to triangulate the quantitative results.

Ethical Considerations - In adherence to the ethical principles of the American Psychological Association (2017) and the Philippine Data Privacy Act of 2012 (Republic Act No. 10173), rigorous human-subjects and data protection protocols were enforced across data collection, analysis, storage, and disposal. Prior to survey administration, formal written authorization was secured from the Campus Administrator and concerned College of Education faculty at CBSUA–Pasacao. Participation was strictly voluntary, and all respondents executed an electronic informed consent form detailing the study's purpose, absence of academic or grade incentives, and unrestricted right to withdraw at any stage without penalty. Beyond informed consent, four technical data protection safeguards were implemented during data processing and storage: (a) *Data Minimization and De-Identification*—Google Forms settings were configured to disable automatic institutional email harvesting, optional participant names were immediately purged upon spreadsheet export, and each record was assigned an anonymized alphanumeric code (*R001* to *R158*) prior to statistical computation and thematic coding; (b) *Encrypted Storage and Access Control*—all raw and coded datasets were stored exclusively in a password-protected, two-factor-authenticated institutional cloud repository and an encrypted local directory accessible solely to the six researchers and their faculty adviser; (c) *AI Data Privacy Firewall*—no raw student survey records, open-ended verbatim transcripts, or personally identifiable metadata were uploaded into commercial generative AI platforms at any stage of analysis; and (d) *Data Retention and Disposal*—the de-identified analytical dataset will be retained securely for a maximum of three years following journal publication to support academic verification, after which all digital files will be permanently overwritten and purged.

3. Results and Discussion

This section presents the descriptive, correlational, and thematic findings derived from the 158 first-year to third-year BSED Science students at Central Bicol State University of Agriculture–Pasacao Campus, organized sequentially according to the four research questions alongside their theoretical and empirical interpretation.

3.1 Extent of AI Chatbot Use as a Learning Aid

The first research question examined the extent of AI chatbot use as a learning aid among first-year to third-year BSED Science students across three functional dimensions: concept explanation (Table 2), problem-solving support (Table 3), and review and reinforcement of scientific concepts (Table 4), with the dimensional synthesis presented in Table 5.

Table 2
Extent of AI Chatbot Use in Terms of Concept Explanation

Indicators	M	Rank	Interpretation
I use AI chatbots to explain difficult scientific concepts.	4.31	1	Very High
I use AI chatbots to clarify topics I do not understand.	4.27	2	Very High
I use AI chatbots to provide examples.	4.22	3	Very High
I use AI chatbots to re-explain lessons step-by-step.	4.16	4	High
I use AI chatbots to simplify complex lessons.	4.13	5	High
Overall Mean	4.22	—	Very High

Note. N = 158. Indicators are ordered from highest (Rank 1) to lowest (Rank 5) weighted mean. Scale: 4.21–5.00 = Very High; 3.41–4.20 = High; 2.61–3.40 = Moderate; 1.81–2.60 = Low; 1.00–1.80 = Very Low.

As shown in Table 2, concept explanation achieved an overall weighted mean of M = 4.22 (Very High). Respondents relied most heavily on AI chatbots to explain difficult scientific concepts (M = 4.31, Rank 1), clarify unfamiliar topics (M = 4.27, Rank 2), and provide illustrative examples (M = 4.22, Rank 3), while step-by-step lesson re-explanation (M = 4.16, Rank 4) and simplifying entire complex lessons (M = 4.13, Rank 5) rated slightly lower within the High range. This pattern indicates that BSED Science students turn to conversational AI primarily when encountering dense terminology or abstract principles requiring immediate clarification, while still relying on faculty instruction and textbooks to synthesize broader curricular units. Theoretically, this aligns with Vygotsky's (1978) Zone of Proximal Development, wherein conversational AI functions as an on-demand dialogic scaffold that bridges initial conceptual gaps, and corroborates higher education syntheses identifying concept clarification as the primary driver of student chatbot engagement (Deng et al., 2025; Kasneci et al., 2023).

Table 3
Extent of AI Chatbot Use in Terms of Problem-Solving Support

Indicators	M	Rank	Interpretation
I use AI chatbots to understand solution steps.	4.12	1	High
I use AI chatbots to analyze scientific questions.	4.06	2	High
I use AI chatbots to guide me in solving problems.	4.04	3	High
I use AI chatbots when I get stuck.	4.00	4	High
I use AI chatbots to check if my answers are correct.	3.97	5	High
Overall Mean	4.04	—	High

Note. N = 158. Indicators are ordered from highest (Rank 1) to lowest (Rank 5) weighted mean. Scale: 4.21–5.00 = Very High; 3.41–4.20 = High; 2.61–3.40 = Moderate; 1.81–2.60 = Low; 1.00–1.80 = Very Low.

Table 3 shows that problem-solving support yielded an overall weighted mean of M = 4.04 (High). Notably, respondents prioritized procedural scaffolding—using chatbots to understand solution steps (M = 4.12, Rank 1), analyze scientific questions (M = 4.06, Rank 2), and guide problem-solving processes (M = 4.04, Rank 3)—over merely checking whether final answers were correct (M = 3.97, Rank 5). Through the lens of Cognitive Load Theory (Sweller, 1988; Sweller et al., 2011), worked examples and segmented procedural guidance reduce extraneous cognitive load during complex quantitative or analytical tasks, allowing learners to allocate working memory resources toward underlying scientific rules (Gkintoni et al., 2025). This finding parallels experimental evidence by Urban et al. (2024), who demonstrated that ChatGPT-assisted scaffolding enhances university students' problem-solving efficacy when learners attend actively to solution logic.

Table 4
Extent of AI Chatbot Use in Terms of Review and Reinforcement

Indicators	M	Rank	Interpretation
I use AI chatbots to summarize key concepts.	3.96	1	High
I use AI chatbots to generate practice questions.	3.95	2	High
I use AI chatbots to reinforce learning.	3.83	3	High
I use AI chatbots to recall learned information.	3.80	4	High
I use AI chatbots to review lessons before exams.	3.72	5	High
Overall Mean	3.85	—	High

Note. N = 158. Indicators are ordered from highest (Rank 1) to lowest (Rank 5) weighted mean. Scale: 4.21–5.00 = Very High; 3.41–4.20 = High; 2.61–3.40 = Moderate; 1.81–2.60 = Low; 1.00–1.80 = Very Low.

As reported in Table 4, review and reinforcement of scientific concepts obtained an overall weighted mean of M = 3.85 (High). Respondents utilized AI chatbots most frequently to summarize key concepts (M = 3.96, Rank

1) and generate practice questions (M = 3.95, Rank 2), whereas reviewing lessons directly before examinations ranked lowest across the entire utilization scale (M = 3.72, Rank 5). During high-stakes examination preparation, preservice science teachers likely prioritize instructor-validated lecture slides, laboratory notes, and textbooks to avoid studying potentially imprecise AI-generated details. This selective pattern aligns with science education studies showing that chatbots function effectively as supplementary review and self-regulation tools alongside core course materials (Deveci Topal et al., 2021; Prondoza & Panoy, 2022).

Table 5
Summary of the Extent of AI Chatbot Use as a Learning Aid

Functional Dimensions	M	Rank	Interpretation
Concept Explanation	4.22	1	Very High
Problem-Solving Support	4.04	2	High
Review and Reinforcement	3.85	3	High
Overall AI Chatbot Use	4.04	—	High

Note. N = 158. Rank 1 indicates the highest dimensional mean. Scale: 4.21–5.00 = Very High; 3.41–4.20 = High; 2.61–3.40 = Moderate; 1.81–2.60 = Low; 1.00–1.80 = Very Low.

Table 5 synthesizes the overall extent of AI chatbot use across the three functional dimensions, yielding a composite mean of M = 4.04 (High). Concept explanation ranked first (M = 4.22, Very High), followed by problem-solving support (M = 4.04, High) and review and reinforcement (M = 3.85, High). Consistent with the Technology Acceptance Model (Davis, 1986), this hierarchy demonstrates that students perceive conversational AI as most immediately useful when decoding unfamiliar scientific concepts, moderately high when navigating procedural problem sets, and supplementary when consolidating material for long-term review.

3.2 Level of Perceived Learning Comprehension

The second research question assessed the respondents' level of perceived learning comprehension across three cognitive dimensions: understanding of scientific concepts (Table 6), retention of information (Table 7), and application of scientific concepts (Table 8), with the dimensional synthesis presented in Table 9.

Table 6
Level of Perceived Learning Comprehension in Terms of Understanding of Scientific Concepts

Indicators	M	Rank	Interpretation
I understand relationships between ideas.	3.99	1	High
I can identify key ideas.	3.89	2	High
I can explain concepts in my own words.	3.85	3	High
I can interpret scientific explanations.	3.75	4	High
I can explain scientific concepts clearly.	3.72	5	High
Overall Mean	3.84	—	High

Note. N = 158. Indicators are ordered from highest (Rank 1) to lowest (Rank 5) weighted mean. Scale: 4.21–5.00 = Very High; 3.41–4.20 = High; 2.61–3.40 = Moderate; 1.81–2.60 = Low; 1.00–1.80 = Very Low.

As shown in Table 6, perceived understanding of scientific concepts achieved an overall weighted mean of M = 3.84 (High). Respondents reported the highest confidence in understanding relationships between scientific ideas (M = 3.99, Rank 1) and identifying key ideas (M = 3.89, Rank 2), but comparatively lower confidence in interpreting scientific explanations (M = 3.75, Rank 4) and explaining scientific concepts clearly (M = 3.72, Rank 5). In constructivist learning theory (Greeno, 1997; Vygotsky, 1978), receptive comprehension (recognizing conceptual links when explained by an instructor or chatbot) precedes productive mastery (articulating mechanisms independently using precise scientific discourse), highlighting the need for classroom tasks that require preservice teachers to verbalize and defend their explanations.

Table 7 indicates that perceived retention of information yielded an overall weighted mean of M = 3.82 (High). While respondents felt confident recalling lessons during examinations (M = 3.92, Rank 1) and remembering key scientific terms (M = 3.89, Rank 2), retaining scientific information over extended periods ranked lowest (M = 3.72, Rank 5). This gap between short-term exam recall and durable long-term retention mirrors Table 4's finding that students underutilize chatbots for systematic review. Under Cognitive Load Theory (Sweller et al., 2011),

durable schema consolidation in long-term memory requires active retrieval practice and spaced elaboration rather than passive reading of AI-generated summaries; without deliberate retrieval effort, rapid AI assistance can create an illusion of competence that fades after immediate assessments (Urban et al., 2024).

Table 7
Level of Perceived Learning Comprehension in Terms of Retention of Information

Indicators	M	Rank	Interpretation
I can recall lessons during exams.	3.92	1	High
I remember key terms.	3.89	2	High
I recall scientific processes.	3.78	3	High
I can remember scientific concepts.	3.77	4	High
I retain information over time.	3.72	5	High
Overall Mean	3.82	—	High

Note. N = 158. Indicators are ordered from highest (Rank 1) to lowest (Rank 5) weighted mean. Scale: 4.21–5.00 = Very High; 3.41–4.20 = High; 2.61–3.40 = Moderate; 1.81–2.60 = Low; 1.00–1.80 = Very Low.

Table 8
Level of Perceived Learning Comprehension in Terms of Application of Scientific Concepts

Indicators	M	Rank	Interpretation
I can apply learning to new situations.	3.94	1	High
I can apply concepts to real-life problems.	3.92	2	High
I can analyze situations using science.	3.84	3	High
I can answer analytical questions.	3.73	4	High
I can solve problems independently.	3.72	5	High
Overall Mean	3.83	—	High

Note. N = 158. Indicators are ordered from highest (Rank 1) to lowest (Rank 5) weighted mean. Scale: 4.21–5.00 = Very High; 3.41–4.20 = High; 2.61–3.40 = Moderate; 1.81–2.60 = Low; 1.00–1.80 = Very Low.

As displayed in Table 8, perceived application of scientific concepts obtained an overall weighted mean of M = 3.83 (High). Respondents reported higher confidence when relating concepts to new situations (M = 3.94, Rank 1) and real-life problems (M = 3.92, Rank 2), but noticeably lower assurance when answering analytical questions (M = 3.73, Rank 4) and solving scientific problems independently (M = 3.72, Rank 5). In Vygotskian scaffolding theory (Vygotsky, 1978), effective instructional support must gradually fade as learner competence matures; if students rely continuously on conversational AI hints during problem sets, their autonomous problem-solving confidence may lag behind their assisted performance (Eteng-Uket & Ezeoguine, 2025).

Table 9
Summary of the Level of Perceived Learning Comprehension

Cognitive Dimensions	M	Rank	Interpretation
Understanding of Scientific Concepts	3.84	1	High
Application of Scientific Concepts	3.83	2	High
Retention of Information	3.82	3	High
Overall Perceived Learning Comprehension	3.83	—	High

Note. N = 158. Rank 1 indicates the highest dimensional mean. Scale: 4.21–5.00 = Very High; 3.41–4.20 = High; 2.61–3.40 = Moderate; 1.81–2.60 = Low; 1.00–1.80 = Very Low.

Table 9 summarizes perceived learning comprehension across the three cognitive dimensions, yielding a composite mean of M = 3.83 (High) across understanding (M = 3.84, Rank 1), application (M = 3.83, Rank 2), and retention (M = 3.82, Rank 3). While the narrow spread across dimensional means reflects a consistent, moderately high level of self-reported comprehension, the shared lowest indicator ratings across Tables 6–8 (M = 3.72 for clear conceptual explanation, long-term retention over time, and independent problem-solving) pinpoint the exact productive skills requiring stronger pedagogical reinforcement.

3.3 Relationship Between AI Chatbot Use and Perceived Learning Comprehension

The third research question tested whether a statistically significant relationship existed between the extent of AI chatbot use as a learning aid and the perceived learning comprehension of first-year to third-year BSED Science students (Table 10).

Table 10

Correlation and Coefficient of Determination Between AI Chatbot Use and Perceived Learning Comprehension (N = 158)

Variables	r	95% CI	r2	Variance (%)	t(156)	p
AI Chatbot Use and Perceived Learning Comprehension	.470	[.338, .584]	.221	22.1	6.651	< .001

Note. Tested using two-tailed Pearson product–moment correlation and *t*-test for correlation significance at α = .05 (*df* = 156). For rounded $r = .470$, $t(156) = 6.651$ (unrounded $r = .4703$ yields $t = 6.657$, $r^2 = .2212$).

As presented in Table 10, bivariate correlation analysis revealed a statistically significant, moderate positive relationship between the extent of AI chatbot use as a learning aid and overall perceived learning comprehension, $r(156) = .470$, 95% CI [.338, .584], $t = 6.651$, $p < .001$. Consequently, the null hypothesis (H_0) was rejected. The coefficient of determination ($r^2 = .221$) indicated that 22.1% of the variance in students' perceived learning comprehension was statistically accounted for by variations in their extent of AI chatbot use, leaving 77.9% of the variance attributable to other instructional, cognitive, and contextual factors.

These inferential results indicate that BSED Science students who engaged more extensively with AI chatbots for concept explanation, procedural problem-solving, and lesson review tended to report higher levels of science learning comprehension, corroborating meta-analytic and experimental evidence on constructive AI scaffolding in higher education (Deng et al., 2025; Urban et al., 2024). Simultaneously, the moderate effect size ($r^2 = .221$) establishes an important pedagogical boundary: nearly four-fifths (77.9%) of the variance in comprehension is governed by factors beyond chatbot utilization—such as quality of faculty instruction, laboratory inquiry, prior scientific knowledge, and autonomous study effort (Eteng-Uket & Ezeoguine, 2025). Furthermore, because comprehension was assessed via self-report in a cross-sectional design, the correlation reflects perceived cognitive benefit rather than verified causal gain, reinforcing that AI chatbots must operate as supplementary scaffolds rather than replacements for core science instruction.

3.4 Thematic Analysis of Open-Ended Responses

To contextualize the quantitative survey ratings, respondents' answers to the three open-ended prompts in Part IV were examined using semantic thematic analysis (Braun & Clarke, 2006) and organized into five overarching themes capturing students' functional uses, verification challenges, and perspectives on responsible AI integration (Table 11).

Table 11

Themes Derived From Open-Ended Responses on AI Chatbot Use (N = 158)

Theme	Analytic Description and Pedagogical Focus
Concept clarification	Simplifying abstract scientific terminology and dense theories through accessible language and concrete examples.
Problem-solving support	Guiding students step-by-step through analytical questions, formula derivations, and procedural solution pathways.
Review and reinforcement	Summarizing lecture notes, consolidating key topics, and generating self-quizzing prompts for formative review.
Accuracy and verification concerns	Recognizing potential inaccuracies, incomplete formulas, or fabricated details in AI outputs and cross-checking against textbooks and lecture slides.
Responsible use of AI chatbots	Utilizing chatbots strictly as supplementary learning scaffolds rather than substitutes for independent reasoning, laboratory inquiry, or teacher instruction.

Note. Themes were derived through semantic thematic analysis (Braun & Clarke, 2006) of open-ended responses from Part IV of the questionnaire (*N* = 158).

As outlined in Table 11, the first three qualitative themes—concept clarification, problem-solving support, and review and reinforcement—triangulate the quantitative hierarchy in Table 5, confirming that students turn to chatbots primarily to translate abstract science terminology into accessible language, unpack step-by-step problem solutions, and condense lecture notes. Crucially, the fourth and fifth themes—accuracy and verification concerns, and responsible use of AI chatbots—illuminate the mechanisms behind the quantitative ratings. Respondents noted that chatbots occasionally generate imprecise scientific explanations or incomplete chemical and physical formulas,

prompting conscientious learners to cross-check outputs against textbooks and instructor slides. Respondents further emphasized that uncritical copying undermines independent problem-solving during proctored examinations, reinforcing the necessity of structured instructional guardrails.

3.5 Pedagogical Framework and Instructional Recommendations

Addressing the fourth research question, the empirical profiles (Tables 2–9), bivariate correlation (Table 10), and qualitative themes (Table 11) were synthesized into an evidence-based pedagogical framework (Table 12) to support effective, ethical, and responsible AI chatbot integration among BSED Science students.

Table 12
Pedagogical Framework and Instructional Recommendations for AI Chatbot Integration

Empirical Finding	Instructional Recommendation	Theoretical Alignment
Concept explanation ranked highest (M = 4.22, Very High)	Use chatbots for initial concept clarification and examples while requiring mandatory verification against peer-reviewed textbooks and lecture materials.	Cognitive Load Theory (Gkintoni et al., 2025; Sweller, 1988; Sweller et al., 2011); TAM (Davis, 1986)
Problem-solving support rated High (M = 4.04) but independent problem-solving rated lowest (M = 3.72)	Position chatbots as faded step-by-step scaffolds while requiring students to submit independent reasoning and closed-book problem solutions.	Social Constructivism & ZPD (Greeno, 1997; Van Meter & Stevens, 2000; Vygotsky, 1978)
Review and reinforcement ranked lowest (M = 3.85) alongside long-term retention (M = 3.72)	Design structured AI prompts for active retrieval practice, self-quizzing, and spaced elaboration rather than passive summary reading.	ADDIE Instructional Design Model (Branch, 2009; Molenda, 2003); CLT (Sweller et al., 2011)
High perceived comprehension across understanding (M = 3.84), application (M = 3.83), and retention (M = 3.82)	Blend AI scaffolding with teacher-led discourse, hands-on laboratory inquiry, peer collaboration, and oral conceptual defense.	TPACK Framework (Koehler & Mishra, 2009; Mishra & Koehler, 2006)
Moderate positive correlation with comprehension (r = .470, r2 = .221) and accuracy concerns	Establish explicit course policies on ethical AI disclosure, source verification rubrics, and boundaries against passive answer copying.	Activity Theory AIED Ecosystem & Responsible AI Governance (Alvarez et al., 2024; Dwivedi et al., 2023; Gile & Fortes, 2026; Kasneci et al., 2023; Uden, 2024)

Note. Instructional recommendations were synthesized from the descriptive, correlational, and thematic findings of the study (N = 158).

The pedagogical framework in Table 12 capitalizes on students' strongest mode of chatbot engagement—concept explanation (M = 4.22)—while directly targeting the three lowest-rated cognitive indicators identified in the comprehension assessment (M = 3.72 for clear conceptual explanation, long-term retention over time, and independent problem-solving). By pairing initial AI concept clarification with mandatory textbook verification, faded procedural scaffolding (Vygotsky, 1978), spaced retrieval prompts (Branch, 2009; Sweller et al., 2011), and inquiry-based laboratory instruction (Mishra & Koehler, 2006), science teacher educators can harness conversational AI without compromising independent analytical rigor.

3.6 Practical Educational Implications and Significance for Stakeholders

Because the *International Journal of Research Studies in Education* bridges empirical scholarship and classroom practice, the findings of this study yield concrete educational implications for science practitioners and teachers, preservice science students, and higher education institutions:

Significance for Science Practitioners and Teachers. The finding that AI chatbot use accounts for 22.1% of perceived comprehension variance (r = .470) while students rate lowest in explaining concepts clearly (M = 3.72) and solving problems independently (M = 3.72) signals that science educators cannot rely solely on traditional take-home worksheets that chatbots can solve instantaneously. Instead, science practitioners should redesign instructional delivery around *faded scaffolding* and *AI-resilient formative assessment*. Specifically, teachers can assign structured "Critique-the-Chatbot" classroom tasks—in which students prompt an AI chatbot on a complex biology, chemistry, or physics problem and annotate its output against textbook derivations to identify oversimplifications or formula errors (Alvarez et al., 2024)—followed by closed-book problem-solving and oral conceptual defense in the laboratory.

Significance for Students (Preservice Science Teachers). For undergraduate BSED Science majors, the discrepancy between very high chatbot reliance for initial concept explanation (M = 4.22) and lower long-term information retention (M = 3.72) highlights the cognitive danger of mistaking fluent AI explanations for personal mastery (Urban et al., 2024). To convert receptive understanding into durable long-term memory schemas, students should shift from passive summary reading to *active retrieval prompting*—using chatbots to generate diagnostic practice quizzes, attempting analytical problems independently before requesting procedural hints, and paraphrasing scientific mechanisms in their own words. Furthermore, as future secondary school science educators, developing disciplined habits of source verification and transparent AI disclosure prepares these students to model ethical digital literacy for their future K–12 learners (Gile & Fortes, 2026).

Significance for Schools and Teacher Education Institutions. At the institutional level, the findings demonstrate that blanket bans on generative AI are neither realistic nor pedagogically productive, as students already utilize multi-platform AI ecosystems extensively (M = 4.04). Aligning with Uden's (2024) Activity Theory–based AIED ecosystem and Gile and Fortes's (2026) instructional disclosure approach, teacher education institutions and regional state universities should institutionalize clear syllabus guidelines, mandatory AI Use Disclosure templates, and faculty training workshops that embed TPACK-aligned AI scaffolding into science education curricula while safeguarding academic integrity.

4. Conclusions, Limitations, and Recommendations

Conclusions - Based on the empirical and qualitative findings, four main conclusions are drawn. First, first-year to third-year BSED Science students utilize AI chatbots to a high extent as supplementary learning scaffolds, relying on them primarily to decode unfamiliar scientific terminology and trace step-by-step problem procedures rather than for pre-examination review. Second, while respondents report high overall perceived learning comprehension across understanding, application, and retention, their uniformly lower ratings for articulating explanations clearly, retaining information over time, and solving problems independently demonstrate that receptive, AI-assisted understanding does not automatically translate into durable memory consolidation or autonomous problem-solving mastery. Third, purposeful AI chatbot use is significantly and moderately associated with perceived science learning comprehension, accounting for 22.1% of comprehension variance while leaving nearly four-fifths (77.9%) governed by broader instructional, laboratory, and learner factors. Fourth, integrating conversational AI effectively in preservice science teacher education requires structured pedagogical guardrails centering on source verification, faded scaffolding, and active retrieval practice.

Study Limitations and Future Research Directions - Several methodological limitations must be carefully considered when interpreting the empirical findings of this study. First, regarding *cross-sectional design constraints*, all quantitative and qualitative data were collected at a single point in time during Academic Year 2025–2026 using a non-experimental descriptive-correlational design. Consequently, while the bivariate analysis establishes a statistically significant moderate positive covariation between AI chatbot use and perceived learning comprehension (r = .470, r^2 = .221), it cannot establish temporal precedence or directional causality; it remains plausible that students with stronger baseline metacognitive skills utilize AI chatbots more strategically, or that both variables are reciprocally reinforced. Second, regarding *self-report and perceived fluency bias*, both chatbot utilization frequency and learning comprehension were operationalized through structured self-report Likert scales rather than objective, standardized science achievement tests or behavioral prompt logs. Self-reported comprehension captures learners' perceived cognitive mastery and is therefore vulnerable to social desirability bias as well as the metacognitive "illusion of competence," wherein the clarity and immediacy of AI-generated explanations lead students to overestimate their independent retention and analytical readiness (Urban et al., 2024). Third, regarding *sample specificity and institutional delimitation*, the sampling frame was restricted to n = 158 first-year through third-year BSED Science majors at a single Philippine regional state university campus (CBSUA–Pasacao) and excluded fourth-year student teaching interns deployed off-campus. Thus, the findings may not generalize directly to fourth-year clinical interns, non-STEM teacher education programs, or urban universities with different technological infrastructures. Future researchers should employ longitudinal or quasi-

experimental designs that pair self-reported comprehension scales with objective domain-specific science assessments and prompt-log analytics, disaggregate effects across specific LLM platforms (ChatGPT, Google Gemini, Microsoft Copilot), and test the proposed pedagogical framework across multiple teacher education institutions.

Pedagogical and Institutional Recommendations - In light of the findings, preservice science teachers are encouraged to use AI chatbots purposefully for initial concept clarification and active self-quizzing while consistently cross-verifying chatbot outputs against peer-reviewed textbooks and attempting analytical problem sets independently before consulting AI hints. Science faculty instructors should complement AI-assisted study with classroom and laboratory assessments that demand productive cognitive effort—specifically oral and written conceptual defenses, closed-book retrieval practice, and collaborative inquiry without digital aids. At the institutional level, academic administrators within the College of Education at CBSUA–Pasacao should adopt the pedagogical framework in Table 12 to establish transparent syllabus policies on AI source verification, ethical attribution, and responsible AI literacy.

Acknowledgements: The authors express their sincere gratitude to their research adviser, Mr. Deo A. Abiso, as well as the Campus Administrator, College of Education faculty validators, and the 158 first-year to third-year BSED Science students of Central Bicol State University of Agriculture–Pasacao Campus for their invaluable support and participation. Conflict of Interest, Funding, and Publication Fee Confirmation: The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. This undergraduate research received no external funding grants; the authors confirm acceptance of the journal's discounted online publication fee upon formal manuscript acceptance. Data Availability: The anonymized quantitative dataset supporting the findings of this study is available from the corresponding author upon reasonable academic request.

AI Use Disclosure. We used ChatGPT (OpenAI, GPT-4o, March–October 2026), Google Gemini (Google, Gemini 2.5, March–October 2026), QuillBot (v15, 2026), Grammarly (2026), and Liner (2026) for literature organization, initial idea structuring, and language editing during manuscript preparation. All outputs were reviewed, verified, and edited by the authors, including fact-checking primary sources against original journal publications, validating all statistical computations and empirical results, and verifying all theoretical figures and tables. No confidential or personally identifiable data were entered into AI tools. The authors take full responsibility for the content.

Author Contributions Statement (CRediT Taxonomy). In accordance with the Contributor Roles Taxonomy (CRediT; Elsevier & Wiley guidelines, using official Lead/Supporting degree-of-contribution qualifiers), Ar-Jay M. Moya (Lead Researcher & Corresponding Author) led the overall design, execution, statistical and thematic analysis, and manuscript authorship: Conceptualization (Lead), Methodology (Lead), Software & Statistical Analysis (Lead), Validation (Lead), Formal Analysis (Lead), Investigation (Lead), Data Curation (Lead), Visualization (Lead), Writing – Original Draft (Lead), Writing – Review & Editing (Lead), and Project Administration (Lead). The co-authors contributed in supporting capacities as follows: Bernadeth R. Panaligan: Conceptualization (Supporting), Investigation (Equal), Validation (Equal), Writing – Review & Editing (Supporting). Aires N. Calugay: Investigation (Equal), Data Curation (Supporting), Resources (Equal), Writing – Review & Editing (Supporting). Ma. Suseth S. Bello: Investigation (Equal), Resources (Equal), Validation (Supporting), Writing – Review & Editing (Supporting). Jerico M. Bonador: Investigation (Equal), Data Curation (Supporting), Validation (Supporting), Writing – Review & Editing (Supporting). Claire C. Aureada: Investigation (Equal), Resources (Equal), Validation (Supporting), Writing – Review & Editing (Supporting). All six authors reviewed and approved the final revised manuscript prior to submission.

5. References

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct* (2002, amended effective June 1, 2010, and January 1, 2017). https://www.apa.org/ethics/code

Alvarez, J. I., Cortez, A. O., & Alberto, M. Z. (2024). Personalized learning in action: Utilizing AI-powered tutors to bridge the gap in mathematics. *International Journal of Research Studies in Education*, *13*(8), 15–25. https://doi.org/10.5861/ijrse.2024.24074

Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer. https://doi.org/10.1007/978-0-387-09506-6

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

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods*

approaches (5th ed.). SAGE Publications.

Davis, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* [Unpublished doctoral dissertation]. Massachusetts Institute of Technology.

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, *227*, Article 105224. https://doi.org/10.1016/j.compedu.2024.105224

Deveci Topal, A., Dilek Eren, C., & Kolburan Geçer, A. (2021). Chatbot application in a 5th grade science course. *Education and Information Technologies*, *26*(5), 6241–6265. https://doi.org/10.1007/s10639-021-10627-8

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Eteng-Uket, S., & Ezeoguine, E. (2025). The impact of artificial intelligence chatbots on student learning: A quasi-experimental analysis of learning outcome and engagement. *Journal of Educators Online*, *22*(2). https://www.thejeo.com/archive/2025_22_2/etenguket__ezeoguine

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Gile, J. M., & Fortes, A. C. G. (2026). Analysis of AI integration strategies of engineering students in ESL composition. *International Journal of Research Studies in Education*, *15*(13), 95–116. https://doi.org/10.5861/ijrse.2026.26224

Gkintoni, E., Antonopoulou, H., Sortwell, A., & Halkiopoulos, C. (2025). Challenging cognitive load theory: The role of educational neuroscience and artificial intelligence in redefining learning efficacy. *Brain Sciences*, *15*(2), Article 203. https://doi.org/10.3390/brainsci15020203

Greeno, J. G. (1997). On claims that answer the wrong questions. *Educational Researcher*, *26*(1), 5–17. https://doi.org/10.3102/0013189X026001005

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeiffer, F., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (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

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education*, *9*(1), 60–70.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Molenda, M. (2003). In search of the elusive ADDIE model. *Performance Improvement*, *42*(5), 34–36. https://doi.org/10.1002/pfi.4930420508

Prondoza, G. R. M., & Panoy, J. F. D. (2022). Development of chatbot supplementary tool in science and the self-regulated learning skills among the Grade 10 students. *Asia Pacific Journal of Advanced Education and Technology*, *1*(Special Issue), 107–116.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

Uden, L. (2024). Activity Theory-based ecosystem for Artificial Intelligence in Education (AIED). *International Journal of Research Studies in Education*, *13*(5), 1–14. https://doi.org/10.5861/ijrse.2024.24000

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study.

Computers & Education, *215*, Article 105031. https://doi.org/10.1016/j.compedu.2024.105031

Van Meter, P., & Stevens, R. J. (2000). The role of theory in the study of peer collaboration. *The Journal of Experimental Education*, *69*(1), 113–142. https://doi.org/10.1080/00220970009600652

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