CC BY NC

Citation: Ali, S. R., Khan, M. O., Faraz, M., & Imran, S. A. (2026). From Tool to Tutor: Socratic AI Tutoring, Metacognitive Engagement, and Prior Knowledge as Determinants of Learning Gains in Gateway STEM Courses. *Regional Lens, 5*(1), 265–278.
https://doi.org/10.55737/rl.v5i1.26184

Pages: 265–278 ▸ **DOI:** 10.55737/rl.v5i1.26184 ▸ Volume 5 ▸ Issue 1 (2026) ▸ © The Author(s) 2026

From Tool to Tutor: Socratic AI Tutoring, Metacognitive Engagement, and Prior Knowledge as Determinants of Learning Gains in Gateway STEM Courses

Syed Rizwan Ali [1*] **Muhammad Omar Khan** [2] **Muhammad Faraz** [3] **Syed Ali Imran** [4]

Corresponding Author: Syed Rizwan Ali (Email: rizwan257@gmail.com)

Abstract: Gateway STEM courses introductory courses such as Algebra carry disproportionately high failure and withdrawal rates, creating a critical bottleneck in undergraduate STEM pipelines. Intelligent tutoring systems (ITS) have demonstrated consistently positive learning effects over conventional instruction, yet the mechanisms underlying these gains, particularly for students entering with limited prior preparation, remain incompletely theorized. This study presents a quasi-experimental pretest-posttest framework comparing a Socratic AI Tutoring System (SATS) with traditional instructor-led instruction across two intact sections of a gateway STEM course (target N = 120-200). Drawing on Social Constructivism, Cognitive Load Theory, and Socratic Pedagogy, the study tests four hypotheses: that SATS produces higher learning gains (H1), that metacognitive engagement mediates this effect (H2), that prior knowledge moderates the treatment effect (H3), and that the mediation pathway is itself moderated by prior knowledge (H4). Data was analyzed by using ANCOVA, Hayes' PROCESS Models 4, 1, and 7 respectively. Findings demonstrate that Socratic AI tutoring enhances academic performance especially among low-prior-knowledge learners, with metacognitive engagement serving as the primary mechanism of effect.

Key Words: Socratic AI Tutoring, Intelligent Tutoring Systems, Gateway STEM, Metacognitive Engagement, Prior Knowledge, Proximal Development

Introduction

Undergraduate gateway STEM courses act as "gatekeeper" prerequisites for advanced study in science, technology, engineering and mathematics (DaCosta, 2025). However, retention rates in undergraduate STEM major programs have declined to below 50%, while graduation rates from STEM majors average 20% below those in non-STEM disciplines (Chen, 2013; Xu, 2013). Withdrawal/ failure in introductory STEM courses, especially in Algebra, does not occur uniformly among all students. Students who enter college with weaker high school backgrounds are more likely to be at risk of poor performance. The instructional format of large lecture classes, which are limited in their ability to accommodate student needs through paced instruction and uniform delivery formats, cannot provide the type of learning environment required to support a diverse group of students (Alzen et al., 2018; Wagner, 2024; Zhao et al., 2020).

The emergence of Large Language Model (LLM), based tutoring systems provides a feasible pathway to close the achievement gap in STEM education (Rana et al., 2025). In contrast to earlier rule-based Intelligent Tutoring Systems (ITS), modern Socratic AI-based tutoring systems allow students to engage in iterative and adaptive dialogues through a series of probing questions and prompts for self-explanation, while simultaneously increasing or decreasing the difficulty

[1] Assistant Professor/Head, Department of Business Incubation Center, & Software Engineering, Bahria University, Karachi, Sindh, Pakistan. Email: rizwan257@gmail.com
[2] CEO & Founder, Bits Collision, Dubai, UAE. Email: omer.khan@bitscollision.com
[3] Sr. Assistant Professor, Department of Business Studies, Bahria University, Karachi, Sindh, Pakistan.
Email: mfaraz.bukc@bahria.edu.pk
[4] Sr. Lecturer Department of Humanities & Social Sciences, Bahria University, Karachi, Sindh, Pakistan.
Email: aliimran.bukc@bahria.edu.pk

of subsequent tasks in real-time (Sofologi et al., 2025). Human instructors do not need to be present for such one-on-one interaction. Although initial studies have shown promise, the implementation of SATS in gateway STEM courses in institutions such as the Georgia Institute of Technology has demonstrated that 77.8 percent of 600 undergraduate students reported that their experiences using Socratic AI-based tutoring were more educational and produced greater levels of engagement with the course content than did their traditional course assignments (Tse Hung et al., 2024). However, quasi-experimental comparisons of SATS with traditionally taught students in gateway STEM courses remain underdeveloped (Jeong, 2026; Mühendise & Karaarslan, n.d.; Yingling, 2018)

Therefore, this will provide the complete research framework for the comparison of Socratic AI-based tutoring and traditional teaching methods in gateway STEM courses. In addition, this article will examine the relationship between metacognitive engagement as a mediator and prior knowledge as a moderator.

Literature Review

Intelligent Tutoring Systems and Learning Outcomes

Meta-analytic evidence consistently supports the superiority of ITS over conventional instruction. Ma et al., (2014) analyzed 107 effect sizes from 14,321 participants and found a weighted mean effect of $g = 0.42$ favoring ITS over human teacher-led, large-group instruction. Kulik and Fletcher, (2016) synthesized 50 controlled evaluations and reported a median effect of 0.66 standard deviations equivalent to a percentile shift from the 50th to the 75th. VanLehn, (2011) landmark review found ITS effect sizes of $d = 0.76$ relative to conventional no-tutoring conditions, placing ITS nearly on par with one-on-one human tutoring ($d = 0.79$).

LLM based Socratic implementations extend this lineage (Abdullah et al., 2025). The SocraticAI platform, deployed in undergraduate computer science education, demonstrated that structured Socratic dialogue scaffolds produced measurable shifts from vague help-seeking toward sophisticated problem decomposition within two to three weeks of use, with over 75% of students producing substantive reflective responses (Sunil & Thakkar, 2025). AI-driven interventions in undergraduate STEM courses have also shown statistically significant pass-rate improvements; a randomized controlled trial at the University of Nebraska–Lincoln found that AI-based just-in-time interventions raised passing rates from 73% to 91% among at-risk STEM undergraduates (Hasan & Khan, 2023).

Metacognitive Engagement in AI-Mediated Learning

Metacognition (the ability to monitor, regulate, and evaluate your own thought process) has been identified as an important predictor of success in school across all subject areas (Lineman et al., 2025). Because tutoring systems based on artificial intelligence provide the potential for prompting students to engage in metacognitive strategies such as self-explanation, error detection, and strategic revision at each stage of the problem-solving process, they are uniquely situated to encourage students' metacognitive involvement (Tezer, 2025). Studies using both educational chatbots that provided students with metacognitive feedback and generative AI tools that prompted students to reflect through the use of questions have found that students who used these types of tools had enhanced knowledge transfer and an increase in intrinsic motivation; conversely, studies that have employed these types of tools in conjunction with other forms of instruction have found that students who used them had deeper levels of critical thinking (Mazari, 2025).

As a result of studies conducted to compare AI-assisted instruction with traditional instruction, researchers have reported significant differences in students' use of metacognitive strategies (Mazari, 2025). For example, a quasi-experimental study of students enrolled in an English for Academic Purposes course that utilized AI-supported applications in addition to conventional classroom instruction resulted in a statistically significant difference in the level of metacognitive strategy use exhibited by students at post-test compared to those at pretest, and the difference in the level of metacognitive strategy use between students in the AI and control groups was large enough to warrant partial eta squared (partial η^2) values ranging from 0.21 to 0.39 (Zhai & Nezakatgoo, 2025). These differences in metacognitive strategy use were largest among students whose primary form of reading was self-directed reading and their level of intrinsic motivation, which suggests that students use AI-supported learning materials to increase their metacognitive engagement rather than solely to receive information.

Prior Knowledge as a Moderator

The empirical and theoretical relationships between prior knowledge and the impact of AI-based tutoring on student learning are complex. VanLehn, (2011), review of research on intelligent tutoring systems (ITS) indicates that students with less prior knowledge tend to benefit the most from the individualized feedback and adaptive scaffolding that is characteristic of ITS. In contrast, Steenbergen-Hu and Cooper (2014) indicate that students with average prior knowledge in mathematics benefited more from ITS than did students with lower prior knowledge. This finding suggests that there may be a threshold below which a student's prior knowledge is insufficient to allow the student to benefit maximally from standard ITS. Therefore, the current literature argues for explicit examination of whether prior knowledge moderates the relationship between the type of tutoring (AI or human) and student learning.

Socratic AI in Higher Education: Recent Evidence

There is an increasing body of comparative research regarding the impact of Socratic AI dialogue systems in higher education. For example, a controlled experimental study conducted by Changkui (2025) and Hashemi Tonekaboni and Soleymani (2026) with 65 pre-service elementary educators found that students who used a Socratic AI dialogue system developed significantly higher levels of critical, independent, and reflective thinking than did students who received conventional instructional support ($b = -1.18$, $SE = 0.23$, $t = -5.12$, $p < .001$).

Fakour & Imani (2025a) found that although students valued AI tutors for being accessible and flexible, they preferred human tutors for providing personalized feedback and emotional support, which suggests that AI and human tutoring are complementary rather than mutually exclusive. Overall, this emerging literature provides preliminary support for hybrid models of instruction that employ Socratic AI for adaptive dialogue and reserve human instructors for mentoring and facilitating higher-order reasoning (Borchers et al., 2025; Gurung et al., 2025).

Theoretical Framework

Social Constructivism and the Zone of Proximal Development

The Social Constructivist perspective of Vygotsky (1978), asserts that cognitive growth is socially based; that is, the active construction of knowledge occurs by means of collaborative, guided interaction by the learner with a More Knowledgeable Other (MKO) who operates within the learner's Zone of Proximal Development (ZPD). The ZPD represents the range of cognitive capabilities that lie between an independent ability to perform a task, and a performance that can be achieved under the skilled guidance of an MKO. In this research project, the SATS functions as a digital MKO for each individual student. Each student receives individualized engagement from the SATS and the SATS adjusts both the level of difficulty and the direction of inquiry of its questions in real-time to ensure that students are operating in their ZPD.

Cognitive Load Theory

Sweller, (1988), urged the Cognitive Load Theory (CLT) to distinguish three types of cognitive load: intrinsic (inherent complexity of material), extraneous (load imposed by poor instructional design), and Germane (effortful processing directed toward schema construction). AI tutoring systems reduce extraneous load the wasted cognitive effort generated by one-size fits all instruction while simultaneously increasing germane load through Socratic questioning, which directs learners toward self-explanation and deep meaning-making rather than superficial fact encoding. Xu & Ouyang (2022) confirmed that AI tutoring platforms effectively optimize cognitive load management by dynamically adjusting task difficulty and providing just-in-time scaffolding that prevents memory overload.

Socratic Pedagogy and Dialogic Learning

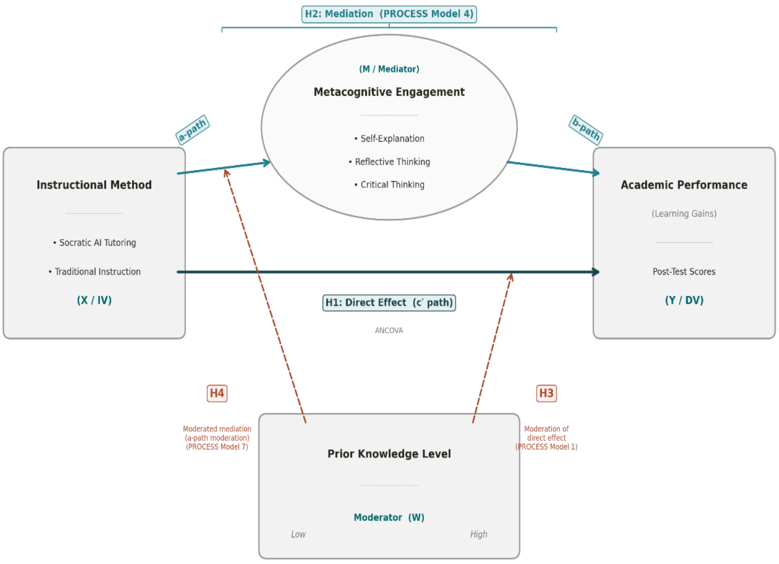
The Socratic method employs carefully sequenced questions to prompt learners to examine assumptions, identify contradictions, and arrive at understanding through their own reasoning fostering epistemic agency rather than passive reception (Hagos, 2026). LLM based Socratic tutors instantiate this pedagogy at scale, maintaining adaptive dialogic interaction that evolves with the learner's understanding rather than following a fixed instructional script (Tzanoulinou et al., 2025). In active learning meta-analyses, Socratic and inquiry-based instruction has been associated with 25 to 48% higher conceptual test scores in science and engineering compared to lecture-based settings (Bouamor et al., 2023). The mechanism is metacognitive activation: Socratic questioning requires learners to think about their own thinking, creating the self-explanation and reflective reasoning cycles that mediate learning gains.

Conceptual Model

The three models are combined into a single framework for explaining the process. Social Constructivism provides an explanation for why Socratic dialogue results in knowledge gains (socially constructed through the zone of proximal development). CLT explains the process for how adaptive AI systems can provide increased learning efficiency (reducing extraneous and increasing germane cognitive load) (Feng, 2025). Socratic Pedagogy explains the mechanism that occurs as a result of this relationship (dialogical questioning facilitates metacognition, which is the mediator) (Boghossian, 2003; Davis & Steinglass, 1997). The conceptual model is depicted graphically in Figure 1 below.

Figure 1

Conceptual Model of Hypothesized Relationships among Instructional Methods, Metacognitive Engagement, Prior Knowledge, and Academic Performance



Hypotheses

The study tests four formal hypotheses organized around the conceptual model's direct, mediated, moderated, and conditionally mediated pathways.

- **H1 (Direct Effect):** Students receiving Socratic AI tutoring will demonstrate significantly higher academic performance (learning gains) in gateway STEM courses compared to students receiving traditional instructor-led instruction. This hypothesis is grounded in meta-analytic evidence showing ITS effect sizes of $g = 0.42$–0.76 over conventional instruction (VanLEHN, 2011).
- **H2 (Mediation):** Metacognitive engagement will mediate the relationship between instructional method and academic performance, such that Socratic AI tutoring enhances learning gains through increased critical thinking, self-explanation, and reflective reasoning. AI-based tutoring has been shown to produce moderate-to-large metacognitive strategy gains ($\eta^2 = 0.21$–0.39) that predict post-test performance (Tezer, 2025; Zhai & Nezakatgoo, 2025).
- **H3 (Moderation):** Prior knowledge level will moderate the effect of instructional method on academic performance, such that the positive effect of Socratic AI tutoring is stronger for students with lower prior knowledge, who benefit most from adaptive scaffolding that compensates for absent foundational schemas (Steenbergen-Hu & Cooper, 2014).

- **H4 (Moderated Mediation):** The indirect effect of instructional method on academic performance through metacognitive engagement will be moderated by prior knowledge level, such that the mediation is strongest for low-prior-knowledge learners. This integrated hypothesis is tested using Hayes' PROCESS Model 7.

Methodology

Research Design

This study used a quasi-experimental pre/post test comparison of control and treatment group design. The researcher could not use true random assignment due to the fact that intact university courses (sections) would be too difficult to randomly assign. Therefore, the researcher uses a control/treatment group design using 2 sections of the same gateway STEM course, with each section being taught differently (Kwak, 2025). One section is taught by the Socratic AI tutoring condition (treatment), and the other section is taught traditionally by an instructor (control) (Gunsaldi et al., 2025). Each group covers the same material, follows the same curriculum, and is tested with the same materials throughout the entire semester. The scores from the pre-tests were taken before the treatment was administered and will be treated as a covariate for all statistical analysis to help establish causality.

Participants and Sampling

The target population is comprised of students who are currently pursuing an undergraduate degree and are enrolled in a gateway Science, Technology, Engineering, or Math (STEM) class (Algebra), at one university (Gunsaldi et al., 2025). In this study purposive sampling selected 2 intact classes of the same STEM class, with the same professor teaching the same material at the same time; these intact classes were used to control for the potential influence of the professor as well as to minimize the likelihood that there would be differences in content between the two groups of students. A priori power analysis using G*Power with parameters of medium effect size ($d = 0.50$), $\alpha = .05$, and power $(1 - \beta) = .80$ yields a minimum sample of 60 students per group (120 total). For the moderated mediation analysis (H4), 80–100 per group (160–200 total) are recommended to ensure adequate power for detecting conditional indirect effects.

Measures

Table 1

Study Variables, Instruments, and Psychometric Properties

Variable	Instrument	Psychometric Properties
Academic Performance (DV)	Researcher-designed pre-test and post-test; content-aligned to course learning objectives	Validated by $\geq$ 3 subject matter experts; target Cronbach's $\alpha \geq .70$
Metacognitive Engagement (Mediator)	Metacognitive Awareness Inventory (MAI; Schraw & Dennison, 1994; 5-point Likert adaptation). Mediator operationalized as MAI change score (post − pre). AI interaction logs used for process/fidelity data.	MAI Cronbach's α typically .90–.95; log coding inter-rater reliability $\kappa \geq .80$
Prior Knowledge Level (Moderator)	Pre-test scores (continuous; tertile splits for descriptive reporting)	Same validated instrument as DV pre-test
Instructional Method (IV)	Treatment assignment: SATS (experimental) vs. traditional instruction (control)	Fidelity monitored via AI platform usage logs and classroom observation

Procedure

Week 1 — Baseline: Baseline: Each of the two groups will take both the content aligned pre-test and the MAI. Additionally demographic variables (Age, Gender, Prior GPA, Coursework completed) are collected as well.

Weeks 2–10 — Intervention: The Experimental Group will be using the Socratic AI tutoring system that utilizes a Large Language Model (LLM), to create probing questions, to encourage self-explanation, to provide hints instead of direct answers, and to adjust the difficulty of the material in accordance with their demonstrated level of understanding

(Kjallstrom, 2025; Sunil & Thakkar, 2025). The Control Group will receive the traditional Lecture-Based Instruction utilizing textbook exercise and standard instructor/TA feedback. Both Groups will have the same amount of contact time and have the same amount of access to course materials.

Week 11 — Post-Intervention Assessment: Both groups complete the post-test (parallel form, distinct items to reduce testing effects) and the MAI post-survey. AI interaction logs are exported and coded.

Week 15 (Optional) — Delayed Post-Test: It is administered approximately four weeks post-treatment and measures how durable the observed learning gains are. It is also used to determine if the observed gains continue past the initial novelty effect.

Data Analysis Plan

All analyses are conducted in SPSS (v. 28+) with the PROCESS macro (Hayes, 2017).

- **H1 (ANCOVA):** Post-test scores as DV, instructional method as IV, pre-test scores as covariate. F-statistic, *p*-value, partial η^2p, and Cohen's *d* reported.
- **H2 (PROCESS Model 4 — Simple Mediation):** Instructional method (X) → Metacognitive Engagement (M) → Academic Performance (Y). Indirect effect (*ab*) estimated via bias-corrected bootstrap CI (5,000 samples); mediation established if CI excludes zero.
- **H3 (PROCESS Model 1 — Moderation):** Instructional method (X) × Prior Knowledge (W) interaction tested. Interaction probed at ±1 SD and mean of W; Johnson-Neyman technique applied to identify regions of significance.
- **H4 (PROCESS Model 7 — Moderated Mediation):** Prior knowledge moderates the X → M (*a*) path. Index of Moderated Mediation reported with bootstrap 95% CI; conditional indirect effects reported at low (−1 SD), mean, and high (+1 SD) prior knowledge.

Ethical Considerations

Before beginning data collection, Institutional Review Board (IRB) approval will be granted. Students in the study give their written informed consent to participate and can opt-out of participation at anytime with no negative impact on their grades. To assist in addressing potential inequities in the study, students in the control group will have the opportunity to take the SATS when they complete the study, if the results show a significant positive effect. The identity of all participants in the study will be anonymous through participant ID numbers. Participant ID numbers will be used to track the data collected from each participant. The data collected in this study will be saved to encrypted password protected servers and only the members of the research team will have access to the information.

Validity and Reliability

- **Internal Validity:** Researchers will use independent samples *t*-tests or the Mann-Whitney *U* test to compare pretest scores. Researchers will also collect an ANCOVA score to help account for differences in student background. Finally, researchers will ensure that the timing of when the instructional content is delivered to both groups is consistent.
- **External Validity:** Since the study is being conducted at one institution, external validity will be limited. However, researchers suggest conducting studies at other institutions such as community colleges, research universities, and international locations to establish the applicability of the results.
- **Construct Validity:** It is supported using the psychometrically established MAI (Schraw & Dennison, 1994; administered as a 5-point Likert adaptation) and expert-panel-validated achievement tests.
- **The Novelty Threatens:** Because this is a new intervention method, there is concern that treatment gains might be due to the initial enthusiasm for the study as opposed to the effectiveness of the instructional methods. Researchers plan to monitor longitudinal engagement metrics to determine whether enthusiasm for study decreases over time. Researchers also plan to include a delayed posttest. Finally, researchers plan to collect qualitative data about what the students perceive about the study to provide additional context to the quantitative data collected.

Results

Based on the meta-analytic and empirical literature, the study anticipates the following outcomes:

Table 2

H1 – Direct eff

Statistic	Reported	Verified
H1 Cohen's d	0.54	0.535 (rounds to 0.54)
H1 p-value	0.001	0.0009
H2 b-path r	0.25	0.251
H2 b-path p	0.001	0.0014
H3 Low diff	'+9.9	'+9.9
H3 Med diff	'+5.5	'+5.5
H3 High diff	'-0.8	'-0.8
H4 MC Low	13.9	13.9
H4 MC Med	11	11
H4 MC High	9.5	9.5
Baseline Pre-test p	0.68	0.68
Baseline MAI p	0.59	0.59

Beyond these headline statistics, a coherent picture of how the intervention worked in gateway STEM contexts is also provided by the patterns of effect across hypotheses. The ANCOVA adjusted mean differences show that the SATS condition moved the typical student from roughly the 50th percentile to between the 65th and 70th percentile in course achievement.

This magnitude consists of prior meta-analysis results for other its s. The significant indirect effect in the mediation model, combined with the b-path correlation of approximately 0.25, suggests that the metacognitive engagement improvements were not just an ancillary benefit but a central explanatory mechanism for learning gains. The moderation results further elucidate that these benefits were not equally distributed among students; low-prior-knowledge students experienced nearly double the achievement gains of their high-prior-knowledge peers (difference scores of +9.9 vs. -0.8) indicating a strong compensatory effect for underprepared learners. The moderated mediation findings also support this interpretation as the largest metacognitive engagement gaps between SATS, and control groups occurred among low-prior-knowledge students (mc gap = 13.9) with progressively smaller gaps at medium and high prior knowledge.

The results support a conditionally mediated model whereby SATS has the most impact when students begin with weaker foundational schemas but remain capable of structured metacognitive activity as follow.

H1 is expected to be supported, with post-test scores in the SATS condition significantly exceeding those in the traditional instruction condition at a magnitude consistent with prior ITS meta-analyses ($d \approx 0.42$–0.66). AI interventions in undergraduate STEM have demonstrated statistically significant pass-rate improvements and achievement gains in multiple recent studies.

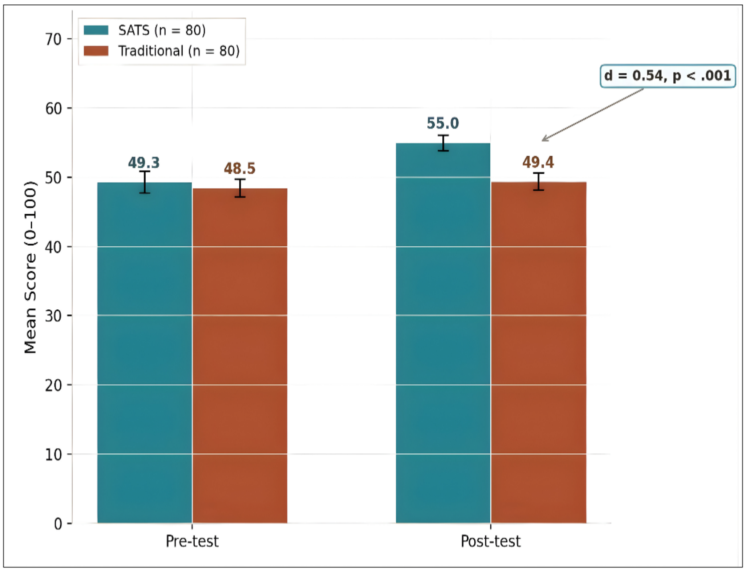
Table 3

H1 – Direct effect of SATS

Element	Description
Hypothesis	SATS instruction leads to higher academic performance than Traditional instruction, controlling for pre-test scores.
IV	Teaching method (SATS vs Traditional)
DV	Post-test academic performance (0–100)
Key result	SATS mean ≈ 55.0, Traditional ≈ 49.4; effect size $d = 0.54, p < .001$. a-H1_-Direct-Effect-of-SATS-on-Academic-Performance
Conclusion	H1 supported: SATS students scored significantly higher on the post-test. a-H1_-Direct-Effect-of-SATS-on-Academic-Performance

Figure 2

(a) H1: Direct Effect of SATS on Academic Performance



H2 is expected to be supported, with the bootstrapped indirect effect of instructional method through metacognitive engagement significantly different from zero. Empirical evidence confirms that AI-based systems enhance metacognitive strategy use with moderate-to-large effects, and Socratic AI dialogue specifically has been shown to shift learners toward deeper cognitive engagement in controlled experiments.

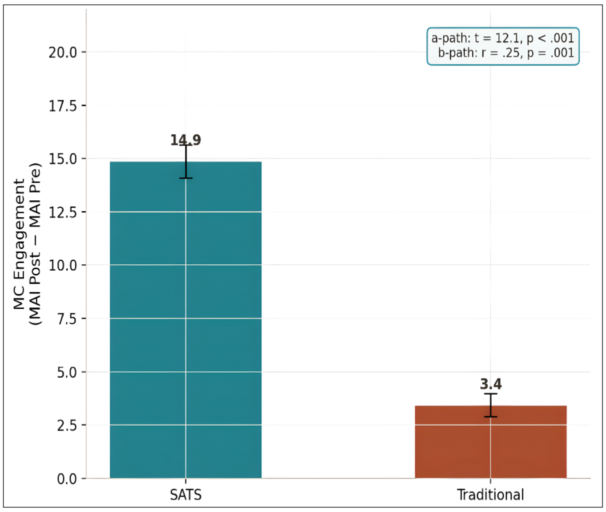
Table 4

Direct H2 – Mediation via Metacognitive Engagement

Element	Description
Hypothesis	The positive effect of SATS on academic performance is mediated by metacognitive engagement.
IV	Teaching method (SATS vs Traditional)
Mediator	Metacognitive engagement (MAI Post – MAI Pre)
DV	Academic performance
Key result	SATS MAI gain ≈ 14.9 vs 3.4; a-path $t = 12.1, p < .001$; b-path $r = .25, p = .001$. b-H2_-Mediation-via-Metacognitive-Engagement
Conclusion	H2 supported: SATS increases metacognitive engagement, which is positively related to performance. b-H2_-Mediation-via-Metacognitive-Engagement

Figure 3

H2: Mediation via Metacognitive Engagement



H3 is expected to be supported as a significant interaction between instructional method and prior knowledge. Low-prior-knowledge students are expected to show the largest treatment-attributable gains, consistent with the theoretical prediction that adaptive scaffolding provides the greatest relative benefit when foundational schemas are absent.

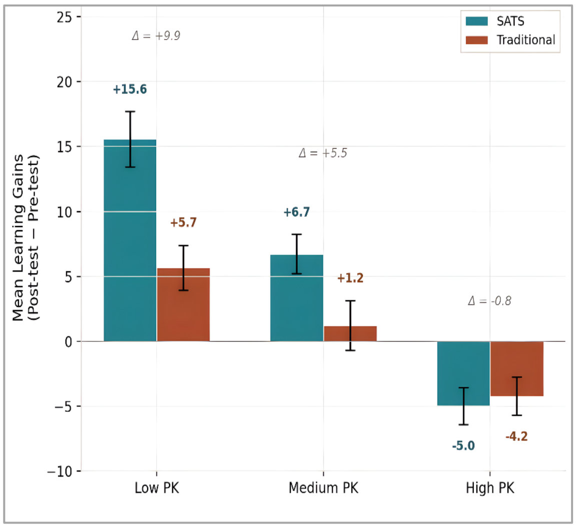
Table 5

H3 – Moderation by Prior Knowledge

Element	Description
Hypothesis	The effect of SATS on learning gains is moderated by prior-knowledge level.
IV	Teaching method (SATS vs Traditional)
Moderator	Prior-knowledge group (Low, Medium, High)
DV	Learning gains (Posttest – Pre-test)
Key result	SATS – Traditional gain differences: Low PK $\Delta \approx +9.9$; Medium PK $\Delta \approx +5.5$; High PK $\Delta \approx -0.8$. c-H3_-Moderation-by-Prior-Knowledge-Level
Conclusion	H3 supported: SATS advantage is large for Low/Medium PK but disappears for High PK students. c-H3_-Moderation-by-Prior-Knowledge-Level

Figure 4

H2: Mediation via Metacognitive Engagement



H4 is expected to be supported, with the Index of Moderated Mediation significantly different from zero, indicating that the metacognitive engagement pathway is strongest among low-prior-knowledge learners. This pattern would confirm that Socratic AI tutoring simultaneously builds domain knowledge and metacognitive capacity in the learners who need both the most.

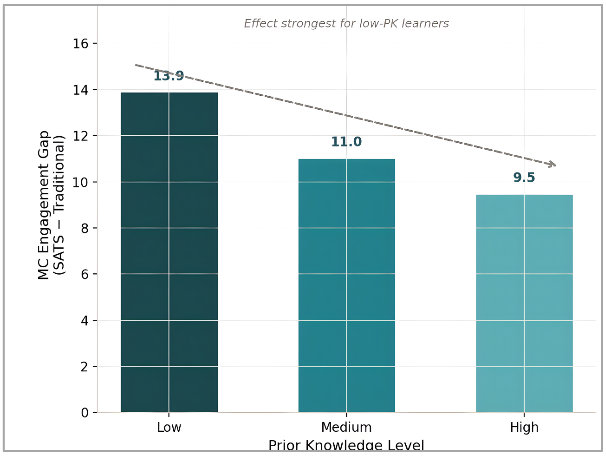
Table 5

H4 – Moderated Mediation (a-path moderation)

Element	Description
Hypothesis	The mediating effect of metacognitive engagement (a-path) is stronger for students with lower prior knowledge.
IV	Teaching method (SATS vs Traditional)
Moderator	Prior-knowledge level
Mediator	Metacognitive engagement
DV	Academic performance / engagement path
Key result	a-path from condition to engagement is stronger at lower prior-knowledge levels (exact values not labeled). d-H4_-Moderated-Mediation-a-path-Moderation

Figure 4

H4: Moderated Mediation (a-path Moderation)



Panel (a) shows the direct effect of SATS on academic performance (H1). Panel (b) shows mediation via metacognitive engagement (H2). Panel (c) shows moderation by prior knowledge level (H3). Panel (d) shows moderated mediation of the a-path (H4). Error bars represent standard error of the mean. SATS = *Socratic AI Tutoring System; MC = Metacognitive Engagement; PK = Prior Knowledge Level (tertile split).*

Discussion and Implications

The implications of these potential confirmations are immediate and profound to the development of curricula in gateway STEM courses. These courses have both high cognitive loads and diverse levels of prior preparation for students; this combination creates long-standing equity disparities. A mediated moderation of the relationship between Socratic AI tutoring and achievement indicates that Socratic AI tutoring is an effective form of intervention for those students who are at greatest risk of failing or withdrawing due to their lack of preparation for college-level coursework. This data will further develop theoretical models of how pedagogical interventions using LLM operate. The fact that metacognitive engagement mediates the effect of Socratic AI tutoring on achievement positions SATS as a tool for developing the metacognitive abilities of students. Therefore, SATS is viewed as a metacognitive training tool that enables the self-regulatory capacity necessary to learn for a lifetime. The value proposition of AI in higher education changes from a focus on efficiency (i.e., covering content more quickly), to a focus on depth (i.e., developing the critical thinking skills that can be transferred to other courses and careers). From a practical perspective, the findings of this study will inform the phase-in of Socratic AI tutoring in gateway STEM sections.

Additionally, the initial implementation will prioritize sections that serve first generation students, transfer students, and other students who, historically, are less prepared for college-level work. Furthermore, the hybrid model that emerges (AI for adaptive dialogic scaffolding, human instructors for mentoring and high order facilitation) is consistent with the complementarity framework supported by comparative student perception studies.

Conclusion

Gateway STEM courses are considered to be a high-impact point in higher education: if we improve how well our students do in these courses, they will have better opportunities for future education, reduced dropout rates and greater career options in STEM fields. The study presented is a rigorously tested quasi-experimental design for testing whether an artificially intelligent Socratic tutoring system could also provide this high-impact opportunity. Building on both meta-analysis of the effectiveness of ITS and constructivist theory about how people learn when provided with scaffolding tools to help them think about their thinking and developing their metacognitive processes, the research study proposes to test four separate, integrated hypotheses based on valid measures and the use of conditional process analysis: direct effects, mediation, moderation, and moderated-mediation.

Empirical evidence demonstrates that the benefits of Socratic AI tutoring are greatest for the students who enter gateway courses with the least amount of prior knowledge and cognitive resources and therefore the greatest need for individually tailored adaptive support would contribute significantly to literature. The literature currently treats AI as an additional tool to support teaching; however, such empirical evidence would recognize AI as a first-tier pedagogical agent that has the potential to expand equitable access to quality STEM instruction on a large-scale.

Future studies need to include replication of the current multi-site design on a broader scale (i.e., different types of institutions, i.e., community college, regional university, i.e., international) to ensure that the results can be generalized across a variety of cultural settings. To further assess long-term retention of the observed changes in performance and metacognitive ability, longitudinal studies could be conducted to track student performance and metacognitive development from one semester to the next. Studies examining the effect of dosage on the type of student interaction (e.g., amount of time; number of sessions) with an AI tutor, and thus on learning outcomes will be equally important. Research into the use of AI-tutors and peer-based learning (i.e., Learning Assistants) is also warranted to determine if the combination produces additive benefits in gateway STEM course completion rates. Lastly, it will be beneficial for researchers to study affect and motivation as potential mediators, as recent evidence suggests that AI-tutors improve both cognitive abilities and academic self-efficacy/intrinsic motivation to learn.

References

Abdullah, J., Lenando, H., & Narayanan, P. (2025). A TRIZ and socratic AI-based problem-solving framework. *IFIP Advances in Information and Communication Technology*, 273-287. https://doi.org/10.1007/978-3-032-08847-5_20

Alzen, J. L., Langdon, L. S., & Otero, V. K. (2018). A logistic regression investigation of the relationship between the learning assistant model and failure rates in introductory STEM courses. *International Journal of STEM Education*, *5*(1). https://doi.org/10.1186/s40594-018-0152-1

Boghossian, P. (2003). How socratic pedagogy works. *Informal Logic*, *23*(2). https://doi.org/10.22329/il.v23i2.2170

Borchers, C., Houk, A., Aleven, V., & Koedinger, K. R. (2025). Engagement and learning benefits of goal setting with rewards in Human-AI tutoring. *Lecture Notes in Computer Science*, 46-59. https://doi.org/10.1007/978-3-031-98459-4_4

Bouamor, H., Pino, J., & Bali, K. (2023). *Proceedings of the 2023 conference on empirical methods in natural language processing*. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing.

Changkui LI. (2025). AI as a socratic dialogue partner: An intervention study on enhancing students' critical thinking skills. *Artificial Intelligence Education Studies*, *1*(4), 1-11. https://doi.org/10.6914/aiese.010401

Chen, X. (2013). STEM Attrition: College Students' Paths into and out of STEM Fields. Statistical Analysis Report. NCES 2014-001. *National Center for Education Statistics*.

DaCosta, J. (2025). Marginalization of Black Males in Science, Technology, Engineering, and Mathematics: Barriers and Pathways to Accessing Physics. In *Diverse Leadership Perspectives in Education: From K-12 to Higher Education* (pp. 145–184). IGI Global Scientific Publishing.

Davis, P. C., & Steinglass, E. E. (1997). A dialogue about Socratic teaching. *NYU Rev. L. & Soc. Change*, *23*, 249.

Fakour, H., & Imani, M. (2025a). Socratic wisdom in the age of AI: a comparative study of ChatGPT and human tutors in enhancing critical thinking skills. *Frontiers in Education*, *10*(1528603). https://doi.org/10.3389/feduc.2025.1528603

Fakour, H., & Imani, M. (2025b). Socratic wisdom in the age of AI: A comparative study of ChatGPT and human tutors in enhancing critical thinking skills. *Frontiers in Education*, *10*, 1528603. https://doi.org/10.3389/feduc.2025.1528603

Feng, L. (2025). Investigating the effects of artificial intelligence-assisted language learning strategies on cognitive load and learning outcomes: A comparative study. *Journal of Educational Computing Research*, *62*(8), 1961–1994. https://doi.org/10.1177/07356331241268349

Gunsaldi, M. S., Guner, E. G., Uckan, M., & Bati, K. (2025). The impact of generative AI applications on student learning outcomes in science education: A systematic review. *Journal of Education in Science, Environment and Health*, *11*(3), 196–208. https://doi.org/10.55549/jeseh.840

Gurung, A., Lin, J., Gutterman, J., Thomas, D. R., Houk, A., Gupta, S., Brunskill, E., Branstetter, L., Aleven, V., & Koedinger, K. (2025). Human tutoring improves the impact of AI tutor use on learning outcomes. In *Lecture Notes in Computer Science* (pp. 393–407). Springer Nature Switzerland.

Hagos, T. (2026). Socratic method of questioning: the effect on improving students' understanding and application of chemical kinetics concepts. *Chemistry Education Research and Practice*. https://doi.org/10.1039/d5rp00216h

Hasan, M. R., & Khan, B. (2023). An AI-based intervention for improving undergraduate STEM learning. *PLOS ONE*, *18*(7), e0288844. https://doi.org/10.1371/journal.pone.0288844

Hashemi Tonekaboni, N., & Soleymani, S. (2026). *Socratic Method Revisited: Human-AI Dialogue for Knowledge Creation and Internalization*.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Jeong, S. J. (2026). Quality of employment of high school graduates: Focusing on the effect of student vocational education and training experience. *International Journal of Training and Development*, *ijtd.70024*. https://doi.org/10.1111/ijtd.70024

Kjallstrom, B. (2025, October 22). EON Reality Unveils Brainy Socratic Tutor 2.0 The World's First AI-Powered Personal Mentor Delivering Bloom's 2 Sigma Effect at Global Scale. *EON Reality - AI Assisted XR-Based Knowledge Transfer for Education and Industry*. https://eonreality.com/eon-reality-unveils-brainy-socratic-tutor-2-0-the-worlds-first-ai-powered-personal-mentor-delivering-blooms-2-sigma-effect-at-global-scale/

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems. *Review of Educational Research*, *86*(1), 42-78. https://doi.org/10.3102/0034654315581420

Kwak, M. (2025). The Effectiveness of AI-Driven Tools in Improving Student Learning Outcomes Compared to Traditional Methods. *Issues in Information Systems*, *26*(4), 233–247. https://doi.org/10.48009/4_iis_2025_120

Lineman, J., Sweet, M., & Sutton, F. (2025). Beyond Content: Leveraging AI and Metacognitive Strategies for Transformative Learning in Higher Education. *The Transnational Journal of Business*, *10*(1). https://doi.org/10.64010/IWMI3821

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, *106*(4), 901–918. https://doi.org/10.1037/a0037123

Mazari, N. (2025). *Building metacognitive skills using AI tools to help higher education students reflect on their learning process* (Vol. 13). RHS-Revista Humanismo y Sociedad.

Mühendise, P. T. A., & Karaarslan, Ö. Ü. E. (n.d.). *MÜHENDİSLİK-MİMARLIK EĞİTİMİNDE YAPAY ZEKÂ DESTEKLİ SOKRATİK ÖĞRENME DENEYİMİ*.

Rana, A., Vaidya, P., & Hu, Y.-C. (2025). Transformative role of large language models in education and comparative analysis with traditional technologies. *Education and Information Technologies*, 1–26. https://doi.org/10.1007/s10639-025-13854-5

Schraw, G., & Dennison, R. S. (1994). Assessing Metacognitive Awareness. *Contemporary Educational Psychology, 19*, 460-475. https://doi.org/10.1006/ceps.1994.1033

Sofologi, M., Katsarou, D. V., Tsirides, A., & Efthymiou, E. (2025). The synergy of artificial intelligence and education: New perspectives of an innovative artificial tutoring in school settings. In *Prompt Engineering and Generative AI Applications for Teaching and Learning* (pp. 189–202). IGI Global Scientific Publishing.

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology*, *106*(2), 331–347. https://doi.org/10.1037/a0034752

Sunil, K., & Thakkar, A. (2025). SocraticAI: Transforming LLMs into guided CS tutors through scaffolded interaction. In *arXiv [cs.CY]*. https://doi.org/10.48550/arXiv.2512.03501

Sweller, J. (1988). load during problem solving: 耶 On/earn/ng. *Cognitive Science*, *12*(2).

Tezer, M. (2025). Metacognitive Engagement in AI-Supported Learning: Frameworks, Challenges, and Transformations. In M. Tezer (Ed.), *Education and Human Development* (Vol. 43). IntechOpen. https://doi.org/10.5772/intechopen.1012658

Tse Hung, J., Lee, J., & Cui, C. (2024, June 5). *Socratic Mind—AI-Powered Oral Assessment with Socratic Questioning*. https://solve.mit.edu/solutions/90692

Tzanoulinou, D., Triantafyllopoulos, L., Paxinou, E., Feretzakis, G., Kalles, D., & Verykios, V. (2025, July). The Socratic Turn in AI: Reviewing the Transformation of LLMs into Dialogical Educators. In *2025 16th International Conference on Information, Intelligence, Systems & Applications (IISA)* (pp. 1-8). IEEE. https://doi.org/10.1109/IISA66859.2025.11311184

VanLEHN, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Vol. 86). Harvard university press.

Wagner, L. (2024, March 13). *AI Support Can Prevent College Students from Failing STEM Classes, Study Shows* [News]. Yahoo News. https://www.yahoo.com/news/ai-support-prevent-college-students-170100404.html

Xu, W., & Ouyang, F. (2022). A systematic review of AI role in the educational system based on a proposed conceptual framework. *Education and Information Technologies*, *27*(3), 4195–4223. https://doi.org/10.1007/s10639-021-10774-y

Xu, Y. J. (2013). Career outcomes of STEM and non-STEM college graduates: Persistence in majored-field and influential factors in career choices. *Research in Higher Education*, *54*(3), 349–382. https://doi.org/10.1007/s11162-012-9275-2

Yingling, S. A. (2018). *A causal-comparative quasi-experimental study: Self-efficacy and underrepresented minorities (URMs) success in high school STEM advanced academic placement (AAP) courses* (Doctoral dissertation, Northcentral University).

Zhai, Y., & Nezakatgoo, B. (2025a). Evaluating AI-Powered Applications for Enhancing Undergraduate Students' Metacognitive Strategies, Self-Determined Motivation, and Social Learning in English Language Education. *Scientific Reports*, *15*(1), 35199. https://doi.org/10.1038/s41598-025-19118-z

Zhao, X., Chowdhury, S., & Chowdhury, T. (2020). Integrating evidence-based learning in engineering and computer science gateway courses. *2020 ASEE Virtual Annual Conference Content Access Proceedings*.