

JOURNAL OF EDUCATION AND HUMANITIES RESEARCH

[✓ OPEN ACCESS](#)

ISSN: 2834-9260 (Online) • Volume 1, Issue 1 (2024) • pp. 37–54

DOI: <https://doi.org/10.5923/researchium.2024.jehr.0603> • **Received:** Sep 12, 2024 • **Accepted:** Oct 08, 2024 • **Published:** Oct 14, 2024 • **License:** CC BY 4.0

Cognitive Load Distribution and Metacognitive Scaffolding in Generative AI-Assisted Undergraduate STEM Classrooms

Dr. Benjamin Foster, Ph.D.^{1*}, Prof. Maria Santos-Cruz, Ed.D.², Dr. Eric Van Houten, Ph.D.²

1. Graduate School of Education, Harvard University, Cambridge, MA, USA

2. Department of Pedagogical Sciences, University of Amsterdam, Amsterdam, Netherlands

*Corresponding Author: bfoster@gse.harvard.edu

ABSTRACT

Background: The integration of Large Language Model (LLM) conversational tutors into higher education STEM classrooms risks creating cognitive offloading and superficial learning. We evaluate an adaptive Socratic GenAI tutoring system across 850 undergraduate engineering students over two academic semesters.

KEYWORDS: Generative AI, STEM Education, Cognitive Load Theory, Metacognition, Educational Technology

1. INTRODUCTION & THEORETICAL CONTEXT

The widespread availability of advanced Large Language Models (LLMs) and conversational generative artificial intelligence (GenAI) agents has disrupted traditional paradigms of higher education, particularly within undergraduate Science, Technology, Engineering, and Mathematics (STEM) disciplines. In complex technical subjects such as multivariable calculus, computational physics, and computer science data structures, students frequently turn to generative AI tools for instant debugging, code generation, and step-by-step problem solutions. While commercial AI interfaces promise personalized 24/7 one-on-one tutoring, their unconstrained deployment raises profound educational and cognitive alarms. According to Sweller's Cognitive Load Theory (CLT), human working memory is severely limited in capacity and duration when processing novel information. Effective learning requires managing three forms of cognitive load: intrinsic load (the inherent complexity of the material), extraneous load (the mental effort imposed by the instructional presentation), and germane load (the productive cognitive effort dedicated to schema construction and conceptual consolidation). When conversational AI systems provide complete, frictionless answers on demand, students engage in cognitive offloading, bypassing productive struggle and failing to build robust mental schemata. Cultivating deep conceptual mastery requires AI pedagogical scaffolding that guides inquiry through Socratic questioning rather than acting as an automated homework-solving oracle.

2. MATERIALS & EXPERIMENTAL METHODOLOGY

We engineered an adaptive Socratic conversational tutoring architecture, termed Socrates-STEM, powered by fine-tuned large language models with instructional prompt constraints that strictly prevent the model from outputting direct solutions or complete code. Instead, Socrates-STEM assesses the student's current conceptual confusion and provides progressive metacognitive scaffolds: clarifying questions, analogous simplified models, hint-based troubleshooting, and self-reflection prompts. We conducted a two-semester randomized controlled pedagogical trial involving 850 undergraduate engineering students enrolled in introductory Object-Oriented Programming and Differential Equations across two large public universities. Students were randomly assigned to one of three conditions: (1) Socratic GenAI Tutor (Socrates-STEM, $n = 285$); (2) Unconstrained Commercial GenAI Tutor (standard ChatGPT-4 access, $n = 285$); and (3) Traditional Teaching Assistant / Textbook Baseline ($n = 280$). We continuously recorded interaction telemetry, eye-tracking pupil dilation (measuring real-time cognitive workload), standardized midterm/final examinations conducted in proctored, AI-free exam halls, and the Metacognitive Awareness Inventory (MAI).

3. RESULTS & EMPIRICAL ANALYSIS

The empirical results demonstrated stark pedagogical divergences between unconstrained AI assistance and Socratic AI scaffolding. In homework assignments, students in the Unconstrained GenAI group completed problem sets 42% faster and earned near-perfect assignment grades (mean 96.4%). However, during proctored, AI-free midterm and final exams, their performance collapsed dramatically: they scored an average of 18.6% lower than the control baseline (mean exam score: 58.2% vs. 71.5%; $t = 7.12$; $p < 0.001$), exhibiting severe conceptual deficits when asked to debug novel algorithms or derive mathematical equations from first principles (the 'GenAI Illusion of Competence'). In sharp contrast, students utilizing the Socratic Socrates-STEM tutor achieved transformative learning gains. Although they spent more time on weekly problem sets than the unconstrained group, their proctored exam scores were significantly higher than both the unconstrained AI and traditional control groups (mean score: 84.8% vs. 71.5% control; $p < 0.001$). Interaction telemetry revealed that Socrates-STEM stimulated high germane cognitive load: students engaged in an average of 4.8 iterative prompt-response refinement cycles per problem, actively diagnosing syntax errors and explaining algorithmic logic. Furthermore, post-course Metacognitive Awareness Inventory scores increased by 31.4% in the Socratic cohort, proving that guided AI tutoring enhances self-regulated learning and problem-solving resilience.

4. DISCUSSION, MECHANISTIC INSIGHTS & LIMITATIONS

The empirical findings illuminate the psychological hazard of frictionless cognitive outsourcing. Unconstrained AI answers induce a dangerous cognitive complacency: students read generated code or solutions, experience a false sensation of fluency, and mistakenly believe they have mastered the material without ever activating the neural retrieval pathways required for long-term schema retention. Productive struggle is not an obstacle to learning; it is the physiological mechanism of learning. Socratic AI scaffolding preserves productive struggle by constraining information delivery, transforming the AI from an answer engine into an intellectual sparring partner. Methodological limitations include the challenge of student resistance: initially, 38% of students in the Socratic condition expressed frustration at the AI's refusal to give direct answers, attempting to 'jailbreak' the system; instructors must explicitly explain the cognitive benefits of Socratic scaffolding at course orientation to secure student buy-in.

5. CONCLUSION & FUTURE TRAJECTORIES

Generative AI possesses tremendous potential to democratize elite-level personalized tutoring in STEM education, but only if engineered with strict pedagogical scaffolding that demands active student cognition. Higher education institutions must reject unconstrained AI answer generators in favor of pedagogically aligned Socratic tutoring systems that build genuine intellectual competence. Future research will explore multimodal AI tutors capable of interpreting student handwritten mathematical derivations in real time.

Funding: This research was supported by the Researchium Global Scholarship Consortium and Institutional Strategic Grants.

Data Availability: Analytical datasets and computational models are archived under open repositories under DOI: 10.5923/researchium.2024.jehr.0603-data.

Conflict of Interest: The authors declare no competing financial or non-financial interests.

REFERENCES

- [1] Bloom BS. The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educ Res.* 1984;13(6):4-16. <https://doi.org/10.3102/0013189X013006004>
- [2] Sweller J. Cognitive Load During Problem Solving: Effects on Learning. *Cogn Sci.* 1988;12(2):257-285. https://doi.org/10.1207/s15516709cog1202_4
- [3] Graesser AC, Person NK, Magliano JP. Collaborative dialogue patterns in naturalistic tutoring. *Cogn Instr.* 1995;13(4):495-522.