Not All Chatbots Teach

Evidence for Pedagogical Design in AI-Assisted Technical Education

Dr. Lucas Cordova Teo Mendoza Kayle Megginson
Ben Webster Sam Holmes Derec Gregory

This is a presentation for a talk for a SIGCITE 2025 conference. As generative AI tools like ChatGPT become embedded in technical education, a critical challenge emerges: how can we ensure these tools foster learning rather than bypass it? This study provides empirical evidence that pedagogical design, not merely model access, determines the educational value of AI assistants. We developed a freely available custom conversational AI tool that embeds metacognitive scaffolding through structured prompts grounded in the Feynman Technique and learning science literature. In a quasi-experimental study within an undergraduate data structures course ($N = 36$), students using this structured AI assistant significantly outperformed peers using the same interface configured as a minimally prompted ChatGPT wrapper (92.7 vs. 74.3, $p < .001$). Gains were especially strong in abstraction, technical justification, and documentation, which are skills critical across software engineering, IT, and cybersecurity. These findings underscore a key insight: AI-integrated learning environments must be intentionally designed to prompt reflection, prediction, and explanation. By aligning AI interactions with evidence-based pedagogy, our framework demonstrates how to develop conceptual understanding, reduce automation bias, and support equitable learning outcomes as AI reshapes computing education.

Table of contents

0.1 Why this paper/talk?

- Generative AI is ubiquitous in CS coursework (79% regular use). [1]
- Unstructured use → trial-and-error, lower self-efficacy, weak transfer. [2], [3], [4]
- Structured, pedagogy-driven prompts → reflection, metacognition gains. [5], [6], [7]

1 Problem & Research Question

1.1 Problem framing

- AI tools can **bypass** core problem-solving steps when unstructured. [4], [8], [9]
- We need to design AI tools to encourage metacognitive skills and reflection, not just provide the answers. [5], [10], [11]

1.2 Research question

> Does integrating metacognitive scaffolding into an AI assistant improve student performance compared to an unstructured Generative AI wrapper with the same model access?

2 Background / Related Work

2.1 Unstructured AI use: risks

- Surface-level engagement; skipping reasoning. [9]
- Lowered self-efficacy, inconsistent outcomes. [2], [3]
- Over-trust in AI recommendations (automation bias in technical tasks, e.g., cybersecurity configs). [12]

2.2 Structured AI design: benefits

- Prompted **explanation/justification/prediction** boosts critical thinking. [13], [14], [14]
- Improves conceptual reasoning for STEM/IT tasks. [15], [16]
- Aligns with ***Explainable AI (xAI)*** and ***self-regulated learning (SRL)***. [6], [11], [17]

3 Pedagogical Tool Design

3.1 Grounded in The Feynman Technique: Explain → Predict → Reflect → Revise

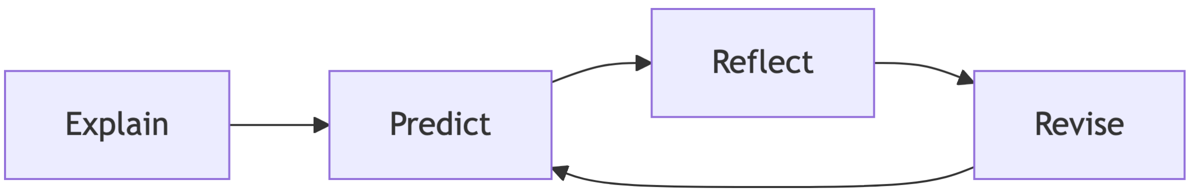


Figure 2: Scaffold operationalizing the Feynman-inspired Explain → Predict → Reflect → Revise loop.

3.2 Pedagogical Learning Theories

- Integrates proven learning theories into the system design to enhance the learning experience:
  - **Feynman Technique** [19]: explain concepts in one's own words to enhance retention and comprehension.
  - **Cognitive Load Theory** [20]: convert verbose analyzer output → concise, relevant guidance (reduce extraneous load).
  - **Zone of Proximal Development** [21]: feedback level aligned to course maturity (scaffolding).

3.3 Pedalogical Application

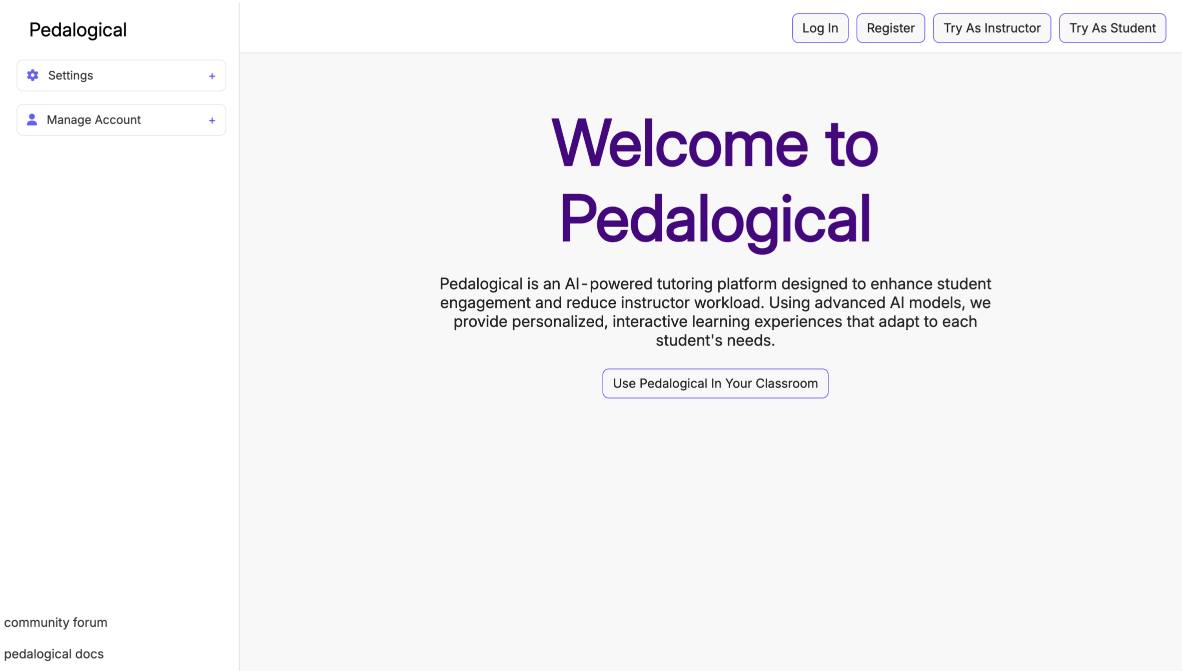


Figure 1: Figure 3: Pedalogical Application

3.4 Pedalogical Question Nodes

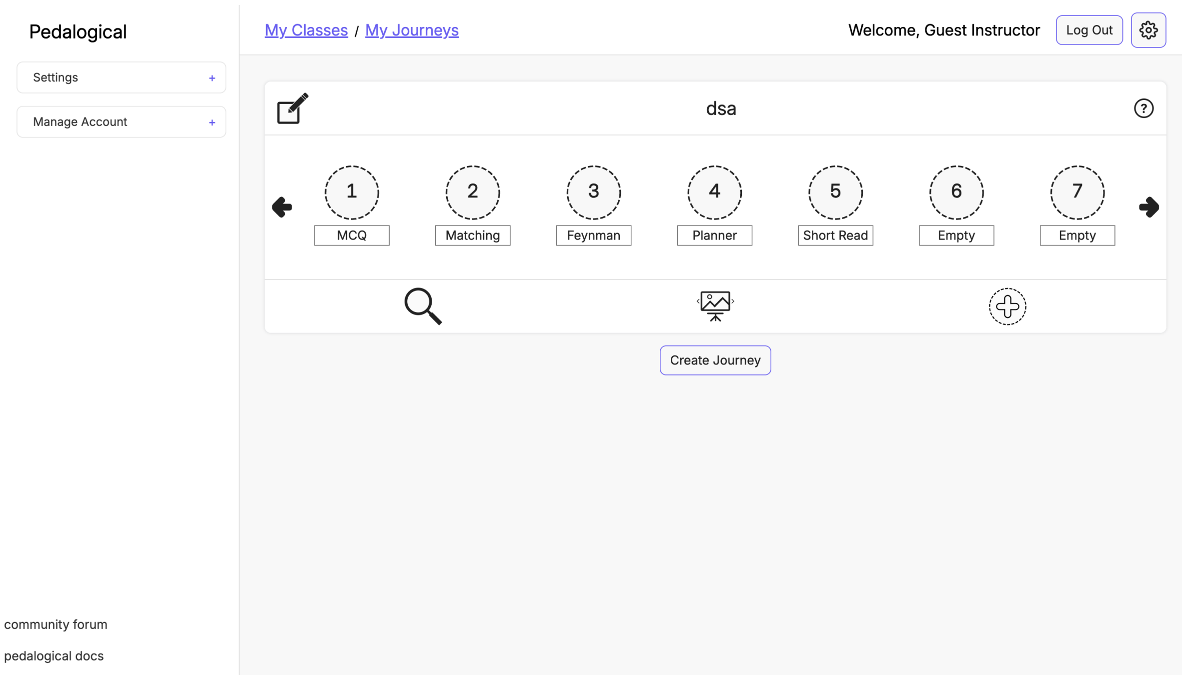


Figure 2: Figure 4: Pedalogical Question Nodes

3.5 Pedalogical LLM Question Generation

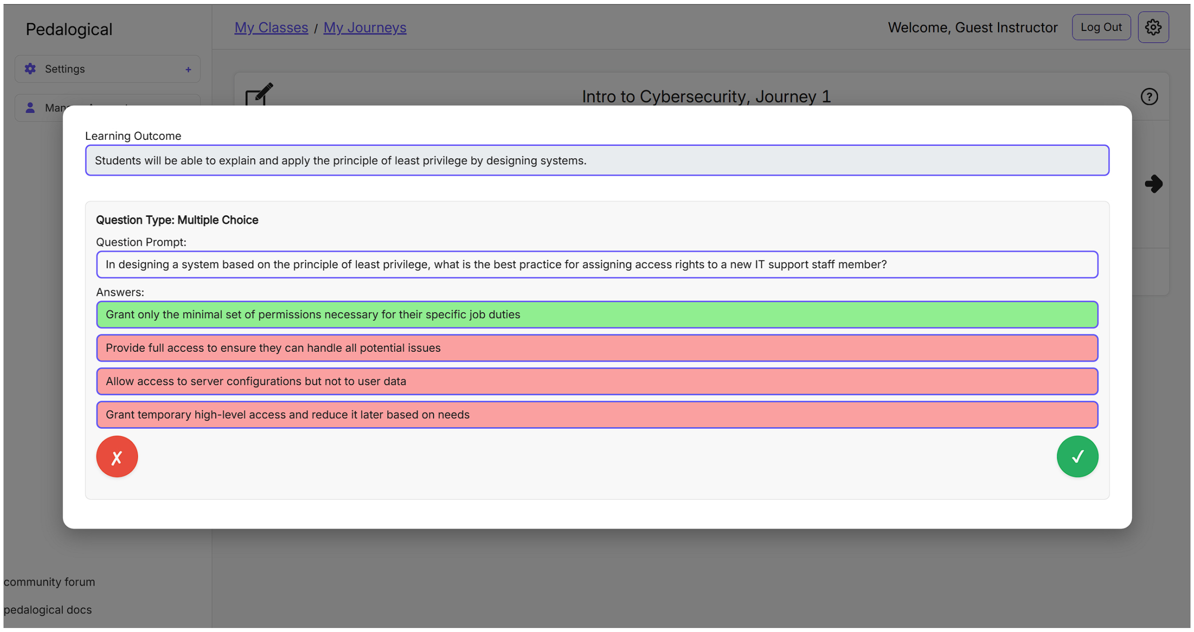


Figure 3: Figure 5: Pedalogical LLM Question Generation

3.6 Prompt intents & light statefulness

- **Concept Articulation** (own words) → clarify mental model.
- **System Reasoning** (constraints) → apply principles.
- **Diagnostic** → surface gaps/assumptions.
- **Justification** → trade-offs and rationale.
- **Light statefulness** → earlier answers steer follow-ups (e.g., check prior claims).

3.7 Beyond programming

- Networking: Subnet reasoning; routing constraints; misconfiguration diagnosis; DNS failure prediction
- Cybersecurity: Risk analysis; firewall rule revision; trade-off justification
- Software Engineering: Debugging; testing rationale; cause–effect reasoning
- Data Science / AI: Model evaluation; metric justification; bias and leakage detection
- Systems Administration: Failure cascade prediction; assumption checking
- Project Management / DevOps: Design trade-offs; pipeline optimization; resource reasoning

4 Study Design

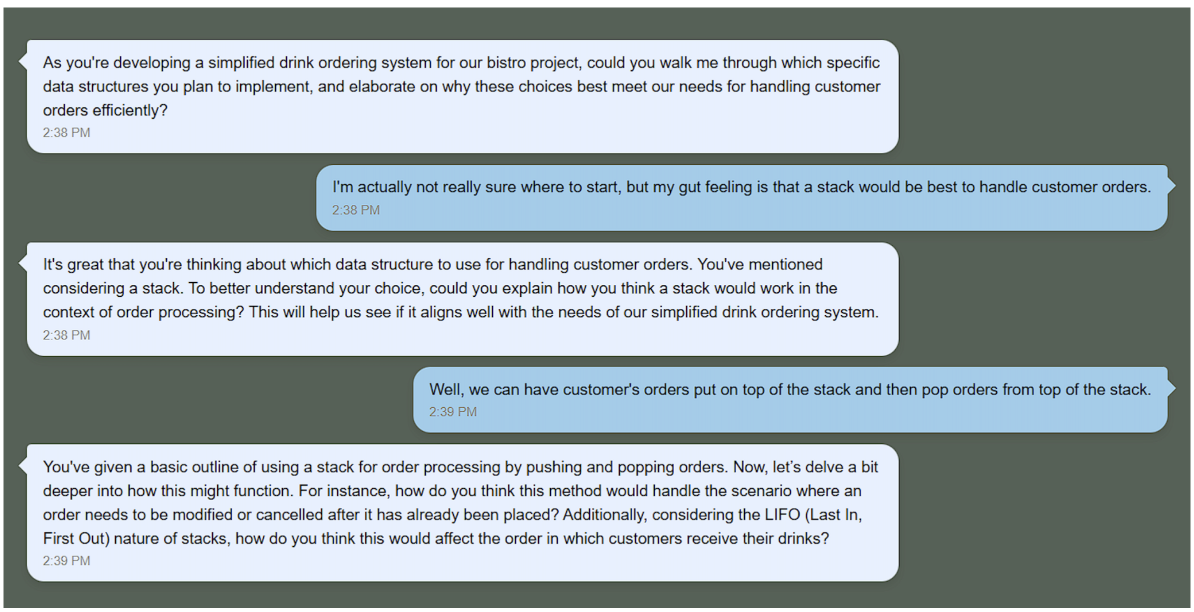
4.1 Context & participants

- Undergraduate Data Structures, Spring 2025, liberal arts university.
- Two sections: **Structured assistant** (n=19) vs. **Unstructured wrapper** (n=17).
- Same instructor, assignments, rubric, incentives.

4.2 Assignment & conditions

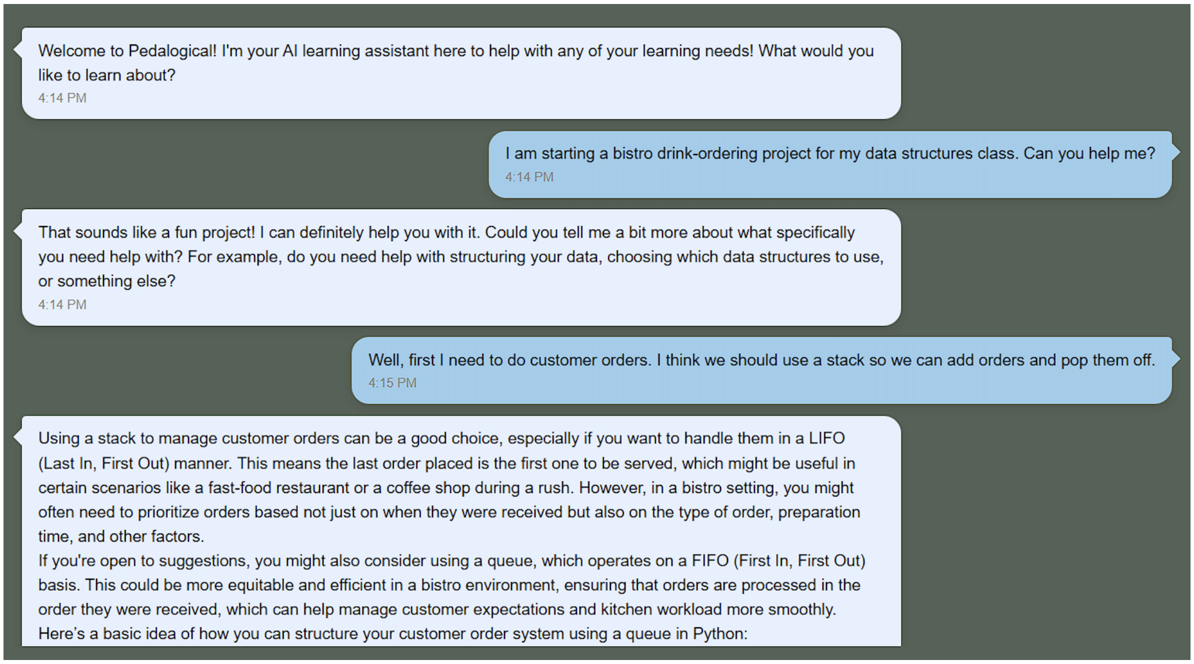
- Project: "Bistro Ordering System."
- Requires data-structure selection/justification, working program, and documentation.
- **Treatment:** gated E-P-R-R prompts with reflection checkpoints.
- **Control:** same UI; minimal pre-prompt; free ChatGPT queries; no scaffolds.

4.3 Treatment: Interaction Snapshot (Structured Assistant)



Conversation scaffold operationalizing the Feynman-inspired Explain → Predict → Reflect → Revise loop.

4.4 Control: Interaction Snapshot (Unstructured Wrapper)



Conversation scaffold operationalizing the Feynman-inspired Explain → Predict → Reflect → Revise loop.

4.5 Measures & analysis

- Blind TA grading; five 20-pt rubric dimensions.
- Overall scores; category scores; logs; short survey.
- Ethics: IRB-approved; de-identified; participation voluntary.

5 Results

5.1 Overall performance

- **Structured**: 92.7 (SD 3.8) vs. **Unstructured**: 74.3 (SD 10.2).
- Independent-samples t-test: $\mathbf{t = 6.93}$, $p < .000001$; **Cohen's d** $=$ **2.14**.
- Lower variance in treatment → more **equitable** outcomes.

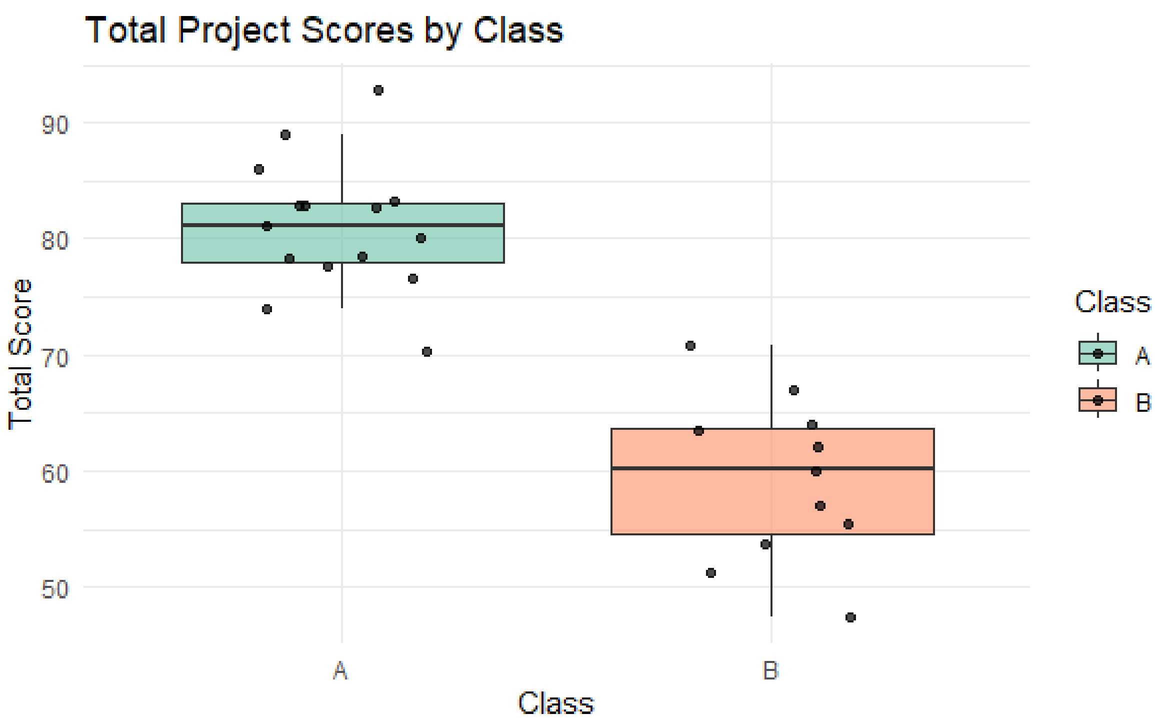
5.2 Where did it help most?

- **Functional implementation** ($p < .0001$)
- **Data structure usage & justification** ($p = .0002$)
- **Report accuracy** ($p = .0005$)
- **Documentation quality** ($p < .0001$)
- No sig. diff. in **code modularity/style** ($p = .27$).

5.3 Interpretation

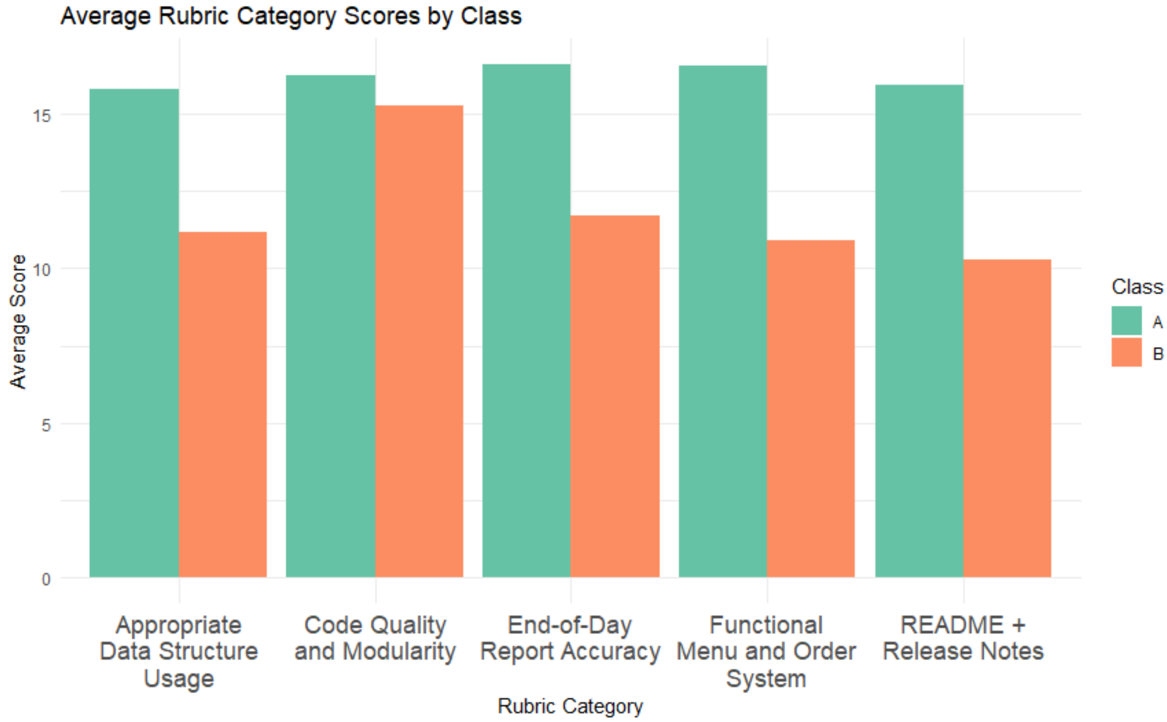
- Pedagogical structure—not just access—drives conceptual gains.
- Scaffolding mitigates **automation bias**; promotes reflective practice. [11, 23]
- Mirrors prior findings on explanation-based learning and metacognition. [1, 16, 23]

5.4 Total Project Scores by Condition



Boxplot illustrating higher average and lower variance for the structured condition. Based on synthetic samples approximating reported summary statistics (92.7±3.8 vs. 74.3±10.2; n=19/17).

5.5 Category Scores by Condition



Schematic relative visualization reflecting significant gains in four rubric categories and no sig. difference in modularity/style. Exact means not reported in paper; shown as proportional (baseline=1.0).

6 Validity & Limitations

6.1 Threats to validity

- Quasi-experimental; section assignment (selection bias).
- Engagement confound (gated prompts vs. free use).
- Product-oriented outcomes; limited metacognitive measures.
- Single institution/course; evolving LLM behavior.

7 Conclusion & Future Work

7.1 Takeaways

- AI's impact is **not pedagogically neutral**. Design matters.
- E-P-R-R scaffolding produced large gains ($d = 2.14$) where it counts.
- Provides a **replicable blueprint** aligned with IT2017/CS2023. [4, 14]

7.2 Where next?

- Randomized, cross-institutional replications.
- Domain-specific scaffolds (cybersecurity, systems, networking).
- Longitudinal learning & transfer; adaptive scaffolds; instructor authoring tools.

7.3 Thank You!

Lucas Cordova

- lpcordova@willamette.edu
- lpcordova.com

Join our lightning talk!

- **11:15 AM - 11:30 AM**: Room **LRC-107**

8 References

8.1 References Cited

[1] J. Finnie-Ansley, J. Becker, M. Denny, P. Luxton-Reilly, B. Simon, and A. Petersen, "Patterns of student use and perceptions of generative AI in advanced computing courses," in *Proceedings of the 56th ACM technical symposium on computer science education (SIGCSE '25)*, New York, NY, USA: ACM, 2025, pp. 123–131. doi: 10.1145/3626252.3627589.

[2] S. Groothuijsen, A. van den Beemt, J. C. Remmers, and L. W. van Leeuwen, "AI chatbots in programming education: Students' use in a scientific computing course and consequences for learning," *Computers and Education: Artificial Intelligence*, vol. 5, p. 100290, 2024, doi: 10.1016/j.caeai.2024.100290.

[3] M. Y. H. Low, C. C. Lee, K. K. Lee, and K. L. Lam, "Enhancing the teaching of data structures and algorithms using AI chatbots," in *2024 IEEE international conference on teaching, assessment and learning for engineering (TALE)*, IEEE, 2024, pp. 1–8. doi: 10.1109/TALE62452.2024.10834293.

[4] X. Zhai *et al.*, "Would ChatGPT-facilitated programming mode impact college students' programming behaviors, performances, and perceptions? An empirical study," *International Journal of Educational Technology in Higher Education*, vol. 21, no. 1, p. 14, 2024, doi: 10.1186/s41239-024-00446-5.

[5] D. J. Liu, J. Markel, and D. J. Malan, "Teaching CS50 with AI: Leveraging generative artificial intelligence in computer science education," in *Proceedings of the 55th ACM technical symposium on computer science education v. 1 (SIGCSE 2024)*, New York, NY, USA: ACM, 2024, pp. 796–802. doi: 10.1145/3626252.3630938.

[6] J. Prather *et al.*, "Metacodenition: Scaffolding the problem-solving process for novice programmers," in *Proceedings of the 25th australasian computing education conference (ACE '23)*, New York, NY, USA: ACM, 2023, pp. 30–39. doi: 10.1145/3576123.3576130.

[7] W. Yan, L. Zhang, W. Xu, and J. Zhou, "Scaffolding computational thinking with ChatGPT," *IEEE Transactions on Learning Technologies*, vol. 17, pp. 1571–1584, 2024, doi: 10.1109/TLT.2024.3392896.

[8] L. Kahn, E. S. Probasco, and R. Kinoshita, "AI safety and automation bias," Center for Security; Emerging Technology, 2024. doi: 10.51593/20230057.

[9] N. Kosmyna *et al.*, "Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task," *arXiv preprint arXiv:2506.08872*, 2025, Available: https://arxiv.org/abs/2506.08872

[10] R. I. A. Ambion, R. S. C. De Leon, A. P. R. Mendoza, and R. M. Navarro, "The utilization of the feynman technique in paired team teaching towards enhancing grade 10 ANHS students' academic achievement in science," in *2020 IEEE integrated STEM education conference (ISEC)*, IEEE, 2020, pp. 1–5. doi: 10.1109/ISEC49744.2020.9397848.

[11] C. Y. Wang, B. L. Gao, and S. J. Chen, "The effects of metacognitive scaffolding of project-based learning environments on students' metacognitive ability and computational thinking," *Education and Information Technologies*, vol. 29, pp. 5485–5508, 2024, doi: 10.1007/s10639-023-12031-z.

[12] K. Wach *et al.*, "The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT," *AI and Ethics*, 2024, doi: 10.1007/s43681-024-00443-4.

[13] C. C. Cao, Z. Ding, J. Lin, and F. Hopfgartner, "AI chatbots as multi-role pedagogical agents: Transforming engagement in CS education," *arXiv preprint arXiv:2308.03992*, 2023.

[14] L. Chen, M. Smith, and R. Johnson, "Scaffolded AI chatbots for enhancing scientific reasoning in undergraduate education," in *Proceedings of the 2025 ACM conference on learning at scale (l@s '25)*, 2025. doi: 10.1145/3591234.3595678.

[15] C. Chhetri and V. Motti, "Exploring large language model-powered pedagogical approaches to cybersecurity education," in *Proceedings of the 25th annual conference on information technology education*, ACM, 2024, pp. 314–319. doi: 10.1145/3686852.3686887.

[16] M. Mukherjee, N. T. Le, Y.-W. Chow, and W. Susilo, "Strategic approaches to cybersecurity learning: A study of educational models and outcomes," *Information*, vol. 15, no. 2, p. 117, 2024, doi: 10.3390/info15020117.

[17] A. Gummadi, K. Santosh, S. S. C. Mary, and B. K. Bala, "Human centric explainable AI for personalized educational chatbots," in *2024 international conference on advanced computing and communication systems (ICACCS)*, IEEE, 2024, pp. 328–334. doi: 10.1109/ICACCS60874.2024.10716907.

[18] S. O. Akinola, B. O. Akinkunmi, and T. S. Alo, "Learn like feynman: Developing and testing an AI-driven feynman bot," in *2024 IEEE international conference on teaching, assessment, and learning for engineering (TALE)*, IEEE, 2024, pp. 1–8. doi: 10.1109/TALE62557.2024.10834370.

[19] M. Haase, "Feynman technique for learning programming and computer science," *Medium: Programming Education Blog*, 2022.

[20] J. Sweller, J. J. van Merriënboer, and F. Paas, "Cognitive architecture and instructional design: 20 years later," *Educational Psychology Review*, vol. 31, no. 2, pp. 261–292, 2019.

[21] L. Vygotsky, *Mind in society: Development of higher psychological processes.* Cambridge: Harvard University Press, 1978.