Evaluation of Generative AI Tools for Lesson Planning in STEM

Rachel Grimes
Graduate Research Assistant
Oklahoma State University
rachel.grimes@okstate.edu

Dr. Kevin Moore
Assistant Professor
Oklahoma State University
kevin.moore@okstate.edu

Introduction/Need for Research

The integration of artificial intelligence (AI) into classrooms has become a transformative force in education, especially within Science, Technology, Engineering and Mathematics (STEM) fields facing challenges like teacher shortages, limited planning time, and burnout (Chiu, 2021; Hall, 2023). Successful implementations of AI, like curriculum evaluation and literacy certification programs, demonstrate its potential in enhancing instruction and workforce readiness (George, 2023; Southworth, 2023), However, barriers persist, including inadequate teacher training and overestimation of AI capabilities (Kong et al., 2021; Kim et al., 2022; Southworth et al., 2023). Educators frequently lack the conceptual knowledge needed to identify which AI tools are best suited to aid in their classroom (Lee & Cho, 2024). Despite these gaps, AI has shown promise in generating instructional resources, assessments, and personalized feedback (George, 2023; Lee et al., 2021). This study contributes to the growing discourse by evaluating the raw capabilities of four AI tools, Claude (Anthropic), Copilot (Microsoft), Gemini (Google), and ChatGPT (OpenAI), in developing course content for STEM educators. The primary aim is to guide platform selection and support effective AI integration in curriculum design (George, 2023; Chen, Chen & Lin, 2020).

Conceptual Framework

To guide this evaluation, the study draws upon the foundations of the TPACK (technological, pedagogical and content knowledge) framework which emphasizes the need for educators to triangulate their teaching approach through technology, pedagogy, and content (Schmidt et al., 2009). TPACK sets the ground for establishing a partnership between the teacher and AI tools to create curriculum (Schmidt et al., 2009). In this process, the teacher provides a direct prompt with content goals and subject to be taught, and the AI tool generates the materials (Schmidt et al., 2009). This framework provides a lens through which the capabilities of AI tools are assessed, particularly in terms of their alignment with educational goals, ethical considerations, and practical classroom integration. Within the TPACK framework, educators must develop technological knowledge specific to AI tools while maintaining their pedagogical expertise and content mastery (Schmidt et al., 2009). Generative AI tools like ChatGPT, Claude, Gemini, and Copilot represent emerging technologies that educators can leverage for curriculum design, administrative efficiency, and instructional planning (Bye, 2018; Krstić et al., 2022). However, successful integration requires educators to understand both the capabilities and limitations of these tools, particularly regarding content accuracy, pedagogical appropriateness, and ethical considerations such as bias and transparency (Lee et al., 2021; Southworth et al., 2023). Many educators lack the training needed for effective AI integration in their curriculum development processes, highlighting the need for restructured teacher education and professional development (Kong et al., 2021; Ng et al., 2022). Despite potential improvements in teaching effectiveness and reduced administrative workload, challenges persist, including data reliability, authentic assessment, and ethical implementation (Rebelo, 2025; Kim et al., 2022; Alam, 2023). While student AI literacy remains important for preparing learners to be responsible users of AI systems (Zhang et al., 2024; Tschoppe et al., 2025), this study focuses specifically on educators' use of AI as a curriculum design tool within the TPACK framework.

Methodology

Four AI tools (Claude, Copilot, Gemini, and ChatGPT) were evaluated for their ability to generate instructional materials for a college-level course covering agricultural safety and health topics. Each bot received a standardized prompt and was asked to produce a complete two-day lesson plan. The prompt provided was: “Create a final lesson plan for a college level safety course that focuses on the safety culture of organizations within the industry. This lesson should last two days which also equates to two 50-minute lectures. Provide a lesson plan with learning goals, objectives, teaching methods and resources. Include a script for the instructor to follow as well as laboratory actives and assignments plans”. This included learning objectives, teaching methods, instructor scripts, lab activities, assignments, grading criteria, and feedback. Following content generation, the bots were tasked with evaluating all lesson plans, including their own, using a standardized rubric. To ensure a comprehensive evaluation, three panels were formed: the AI bots themselves, agricultural education faculty and faculty with content-specific knowledge in agricultural safety and health. This triangulated approach allowed for a balanced assessment of both technical accuracy and pedagogical quality.

Results/Findings

Using the six-dimension rubric developed by the research team (see Methodology), preliminary findings indicate that generative AI tools cannot independently produce course materials with sufficient academic rigor or depth for college-level instruction without substantial educator revision. Across all evaluated dimensions, Claude scored overwhelmingly higher than the other AI tools, with an average of 20.5 out of 24 possible points. The other platforms scored considerably lower: Copilot (12.7/24), Gemini (15.5/24), and ChatGPT (13/24), representing performance gaps of 38-61% compared to Claude.

Conclusions

The findings indicate that while current AI systems may serve as useful tools for idea generation, personalized feedback, and foundational lesson planning, they have not yet achieved the capability to produce complete, presentation-ready lesson plans. The study found that AI does not *yet* (as of the publication date of this study) have the ability to provide detailed lesson materials that align with core objectives and standards better or at the same quality level of the educator. While AI-generated content lacked critical thinking and subject mastery, the tools were effective in generating ideas, offering constructive feedback, and providing a foundation for lesson development. These contributions support the use of AI tools as valuable supplemental resources for educators. In particular, Claude consistently produced the most complete and educator-supportive materials, demonstrating strong potential as a classroom aid.

Implications/Recommendations/Impact

The research team recommends that educators utilize AI tools for tasks less reliant upon detailed and fact-based information. For example, generating bell work questions, identifying exam questions and proving answer options, creating rubrics or lesson outlines is where the strengths of AI tools lay. Additionally, future research is needed in understanding how AI could assist in personalizing learning experiences by analyzing student performance data to tailor educational content to individual needs and providing instant feedback on assignments to enhance the learning process.

References

Bye, R. T. (2018). A flipped classroom approach for teaching a master's course on artificial intelligence. *Communications in Computer and Information Science*, 246–276. https://doi.org/10.1007/978-3-319-94640-5_13

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *Ieee Access*, *8*, 75264-75278.

Chiu, T. K. (2020). Six Key Principles in Designing Artificial Intelligence (AI) Curriculum for Middle Schools. *rdannual*, 383.

Chiu, T. K. (2021). A holistic approach to the design of artificial intelligence (AI) education for K-12 schools. *TechTrends*, *65*(5), 796-807.

George, A. S. (2023). Preparing students for an AI-driven world: Rethinking curriculum and pedagogy in the age of artificial intelligence. *Partners Universal Innovative Research Publication*, *1*(2), 112-136.

Hall, C. J. (2023). *The STEM Teacher Shortage: A Case Study Exploring Arizona Teachers' Job Satisfaction and Retention Intention* (Doctoral dissertation, Northcentral University).

Hajianvari, L., Ramazani, A., (2024). Investigating AI literacy, application and factors influencing artificial intelligence acceptance by faculty members. *Higher Education Letter,* 17. (86): 106-131pages. 10.22034/hel.2024.2036769.1985.

Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. *Education and Information Technologies*, *27*(5), 6069-6104.

Kong, S. C., Cheung, W. M. Y., & Zhang, G. (2022). Evaluating artificial intelligence literacy courses for fostering conceptual learning, literacy and empowerment in university students: Refocusing to conceptual building. *Computers in Human Behavior Reports*, *7*, 100223.

Lee, I., Ali, S., Zhang, H., DiPaola, D., & Breazeal, C. (2021, March). Developing middle school students' AI literacy. In *Proceedings of the 52nd ACM technical symposium on computer science education* (pp. 191-197).

Anthropic. (2025). *Lesson plan for noise-induced hearing loss in agriculture.* [Generative AI Chat] (Sonnet 3.7) https://claude.ai

OpenAI. (2025). *Agriculture safety lesson plan.* [Generative AI Chat]. (GPT5). https://chatgpt.com/

Google. (2025). *Noise-induced hearing loss lesson.* [Generative AI Chat]. (2.5 Flash). https://googlegemini.com

Microsoft. (2025). *Lesson plan for college level safety course in agriculture.* [Generative AI Chat]. (GPT-4-turbo). https://copilot.microsoft.com

Southworth, J., Migliaccio, K., Glover, J., Reed, D., McCarty, C., Brendemuhl, J., & Thomas, A. (2023). Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy. *Computers and Education: Artificial Intelligence*, *4*, 100127.

Stolpe, K., & Hallström, J. (2024). Artificial intelligence literacy for technology education. *Computers and Education Open*, *6*, 100159.