Generative AI in Instructional Design Education: Effects on Novice Microlesson Quality

Steven Moore[1[0000-0002-5256-0339]], Lydia Eckstein[2[0000-0002-9583-7035]], Christine Kwon[1[0009-0001-2825-2280]], and John Stamper[1[0000-0002-2291-1468]]

[1] Carnegie Mellon University, Pittsburgh PA 15213, USA
StevenJamesMoore@gmail.com

Abstract. Instructional designers are increasingly integrating generative AI (genAI) into their workflows to assist with the creation of educational content. However, particularly for novices, it remains unclear whether this integration genuinely elevates the quality of the resulting content. We conducted an A/B field experiment within a 14-week graduate educational technology course (n=27) to examine this potential impact on content quality. Students created eight microlessons across four modules, alternating between receiving genAI assistance (GPT-4 via ChatGPT) and working without genAI. The quality of the microlessons was assessed via a 5-criteria rubric. Results showed that genAI-assisted microlessons received significantly higher quality scores than non-genAI microlessons for half of the assignments and never scored lower on average. Our findings suggest that thoughtful integration of genAI can enhance the quality of materials created by novices.

Keywords: Microlesson, Instructional Design, Generative AI

1 Introduction

Bite-sized learning modules, often called microlessons, are increasingly popular for their alignment with student attention spans and suitability for asynchronous formats [8, 29]. These focused lessons aim to provide targeted instruction, encourage engagement, and support skill mastery [7]. Creating effective microlessons is a key task for instructional designers and learning engineers, yet it often requires substantial time and domain expertise [23]. Generative AI (genAI) is emerging as a transformative force in education [16], offering innovative tools that could potentially ease these content development burdens. Integrating genAI holds promise for supporting designers and even students in the microlesson creation process by assisting with brainstorming, structuring content, and offering feedback.

Despite the promise of genAI, its application in microlesson creation presents notable challenges, especially in educational settings [3, 5]. The quality and pedagogical soundness of AI-generated outputs can be inconsistent, potentially requiring significant user expertise to guide the AI effectively and refine the results [20]. There is also a risk that novice instructional designers may become overly reliant on AI assistance, potentially impeding the crucial development of their own design skills and judgment.

Integrating AI-generated elements seamlessly with learner-developed components also poses a practical hurdle.

Therefore, this study investigates the integration of genAI into the microlesson creation process for novice instructional designers. Conducted within a graduate-level course (n=27) focusing on instructional design and learning engineering, we examined how using genAI influences the quality of student-created microlessons. We implemented a within-subjects A/B experiment where students created eight microlessons throughout the semester, each required to incorporate a distinct learning science principle taught in the course. For alternating microlessons within four course modules, students were assigned to either use genAI for assistance in generating components like learning objectives, assessments, and instructional content, or to complete the microlesson without AI assistance. The quality of all resulting microlessons was systematically evaluated using a pre-defined rubric assessing five key criteria. This research addresses the following primary question:

RQ1: How does the use of genAI by students in a graduate-level course impact the quality of their microlesson designs?

This work offers a key contribution in that it provides empirical evidence demonstrating that incorporating genAI support into the microlesson creation process can lead to significantly higher quality instructional artifacts compared to those developed through traditional, unaided approaches by novice instructional designers.

2 Related Work

Microlessons are brief instructional units (typically 5-15 minutes) focused on several learning objectives [28, 29], featuring aligned content and brief assessments [19]. This structure reduces cognitive load and enhances retention [9, 11], making them widely applicable across educational and professional settings [2, 7, 12]. While educators traditionally create microlessons [18], engaging students in their creation via learnersourcing [14, 21, 26] is an effective pedagogical approach. This process prompts students to research, organize, and present information clearly, deepening their subject comprehension [8, 22] and fostering key skills. Recent advances in genAI tools like ChatGPT offer potential support for this student design process. Prior work found ChatGPT could complete graduate-level instructional design assignments [24], though the quality requires further investigation. Incorporating such tools might assist student designers with brainstorming ideas or refining content, positioning them as active contributors.

3 Methods

This study occurred within a 14-week graduate course covering educational technology and learning science principles during the spring 2024 semester at a large northeastern US university. The course blended bi-weekly in-person sessions (lectures, discussions, activities) with online components completed before class. Participants (n=27) were second-semester master's students in Educational Technology and Applied Learning Science, aged 22-36, reflecting the program's typical demographic profile. As a core

component of their training, students were required to design microlessons throughout the course; this task provided practical experience in applying learning science principles and using educational technology tools, key objectives of the master's program. The intervention focused on four week-long course modules covering Universal Design for Learning (UDL) [25], Guided Discovery (GD) [13], Fostering Help-Seeking (FHS) [27], and Collaborative Learning (CL) [17].

3.1 Learning Platform

The four study modules were delivered via the Open Learning Initiative (OLI) platform [1], an established learning environment featuring instructional content, multimedia, and interactive activities (including formative assessments with feedback). Each module, designed for approximately one week of work, contained 3-6 topics presented across multiple pages with text and videos, interspersed with optional formative activities (multiple-choice, short-answer) offering practice and immediate feedback. Within each module, students were required to create and submit two distinct microlessons totaling eight microlessons per student over the four modules.

3.2 Design

The study followed a within-subjects A/B design [15], where in each of the four modules, students created two microlessons. They were assigned to use ChatGPT (built on GPT-4) for one microlesson (treatment condition) and create the other without AI assistance (control condition). Assignment to the treatment condition for the first versus second microlesson within each module was counterbalanced across the four modules. Consequently, each student produced four AI-assisted and four non-AI-assisted microlessons, enabling direct comparison between conditions while ensuring exposure to both across all module topics.

In the treatment condition microlessons, students were explicitly instructed to use ChatGPT as an assistant throughout the four-step development process described below. Guidance was provided on leveraging the tool for brainstorming ideas, generating initial drafts for learning objectives, assessments, instructional text, and alignment with the designated learning principle. However, specific required prompts or interaction sequences were not prescribed, allowing students flexibility in how they engaged with the genAI assistant within these general guidelines. For the control condition microlessons, students were explicitly instructed not to use genAI tools for any part of the creation process. Adherence to this instruction relied primarily on the university's and course's academic integrity policies and student self-report during submission. We acknowledge the inherent limitations in completely preventing tool usage in a field study setting where direct technical monitoring was not feasible.

3.3 Microlesson Development and Evaluation

Students developed microlessons in the Canvas Learning Management System using a scaffolded four-step process, responding to essay prompts at each stage: 1) Select a topic and identify the required learning science principle for incorporation, 2)

Formulate actionable, measurable, student-centered learning objectives, 3) Create at least three assessments of any format, and 4) Write 1-5 paragraphs of instructional content aligned with the topic and objectives, integrating a strategy related to the designated learning science principle.

Each microlesson had to incorporate a specific learning science principle from the corresponding module, as shown in Table 1. Microlesson quality was assessed using a 5-criteria rubric (Topic Selection, Learning Objectives, Assessments, Instruction, Learning Science Principle Incorporation), provided to students beforehand. Each criterion used a 3-point scale (Exemplary=3, Proficient=2, Needs Improvement=1), yielding a maximum score of 15. The rubric was developed by the instructional team with external feedback and was based on similar rubrics previously used in the literature [4, 10]. Two raters (the course instructor and TA), both with content expertise and microlesson development experience, evaluated the microlessons and were blind to the experimental condition (AI-assisted vs. non-AI-assisted).

Table 1. Overview of the four modules paired with the eight learning science principles used for microlesson development, along with inter-evaluator rubric agreement.

Module Topic	Principle	Description	Kappa (κ)
Universal Design for Learning	Seven Principles	Seven principles of UDL that can be applied as a rubric	0.49
	Eight Steps	Eight steps that demonstrate the process of adopting UDL	0.73
Guided Discovery	Contrasting Cases	Using different examples to highlight nuances between key concepts	0.61
	Tell-then-Practice	Presenting information followed by applied practice opportunities	0.57
Fostering Help-Seeking	Strategies to Improve	Guiding students on strategies to improve their help-seeking behaviors	0.84
	Leveraging Prior Knowledge	Focusing on the connection between prior learning and current challenges	0.82
Collaborative Learning	Cooperative Learning	Structured group work with shared responsibilities	0.82
	Peer Learning	Learning through interaction and feedback between fellow learners	0.79

3.4 Analysis

Prior to independent rating, the two raters participated in training sessions on rubric application and discussed disagreements to consensus on approximately 10% of submissions for each microlesson to ensure consistent interpretation. Inter-rater reliability for the total rubric score on each microlesson was assessed using weighted kappa (κ) [6], accounting for the ordinal nature of the scale. Given the substantial average agreement between raters ($\kappa = .71$, see Table 1 for individual values), the total rubric scores

from both raters were averaged for each microlesson and subsequent analyses utilized these scores.

4 Results

We investigated whether students who used genAI performed differently on each of the eight microlessons compared to students who did not use genAI. Three out of eight microlessons showed a statistically significant advantage for students using genAI, one showed a marginal effect, and four showed no difference. The averaged total rubric scores across each condition for each microlesson can be seen in Figure 1.

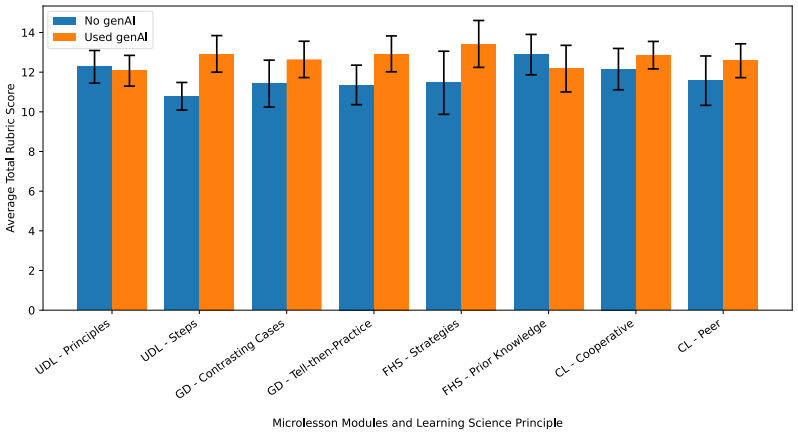

Fig. 1. The total average rubric score for all eight microlessons, divided by the two conditions (genAI use vs. no genAI use), with error bars representing the 95% confidence interval.

Specifically, independent samples t-tests revealed significantly higher quality scores ($p \leq .05$) with genAI assistance for the microlessons on UDL Steps ($t(25) = -4.06$, $p < .001$, Hedges' $g = -1.51$), GD Tell-then-Practice ($t(25) = -2.51$, $p = .02$, Hedges' $g = -0.94$), and FHS Strategies ($t(25) = -2.11$, $p = .05$, Hedges' $g = -0.79$). A marginally significant effect ($p = .09$) favoring genAI was found for GD Contrasting Cases. No significant differences (all $p > .20$) were detected for the UDL Principles, FHS Leveraging Prior Knowledge, CL Cooperative Learning, or CL Peer Learning microlessons. Comparing each student's average score across the four microlessons created with genAI versus the four created without, a paired-samples t-test showed significantly higher scores for the genAI-assisted microlessons ($M = 12.69$, $SD = .99$) compared to the non-genAI ($M = 11.72$, $SD = 1.49$), $t(26) = 4.72$, $p < .001$, Hedges' $g = .88$.

5 Discussion

This study revealed a task-dependent impact of genAI on the quality of microlessons created by novice instructional designers. Our key finding, supported by robust blind

scoring and agreement, showed that genAI assistance significantly enhanced quality for half of the assigned microlessons. It was specifically beneficial for those involving UDL Steps, GD Contrasting Cases, GD Tell-then-Practice, and FHS Strategies, while yielding no significant difference for the remaining tasks and never resulting in lower quality work. This suggests genAI's effectiveness for novices may be greatest on tasks benefiting from structured content generation or rapid information synthesis. Conversely, significant improvements were not observed in this study for tasks potentially requiring more nuanced creativity or complex pedagogical judgment.

The observed quality enhancements in specific tasks align with research suggesting genAI can help novices structure content and avoid common pitfalls [4], potentially streamlining parts of the design process. However, the variability in genAI's impact across different tasks have critical implications for training: novice designers must learn not only how to use genAI, but when to use it strategically and how to critically evaluate and refine its outputs against pedagogical principles and learning goals. Therefore, while genAI presents a valuable tool capable of enhancing the quality of certain novice-created instructional materials, its integration should focus on augmenting human skill and judgment through thoughtful, critical use, rather than replacing the development of core design competencies. For practice, these results suggest genAI tools can be valuable supports for specific tasks, potentially freeing up designer time for more complex pedagogical considerations, but require careful integration and quality control.

While conducting this A/B experiment as a field study within an authentic course offers ecological validity, certain limitations warrant consideration. Primarily, the study's scope being a single course with a limited number of participants (n=27) means caution is needed when generalizing findings to broader contexts or different learner groups. Furthermore, the classroom setting introduces the inherent challenge of ensuring complete treatment fidelity; we cannot entirely rule out the possibility that students in the control condition (instructed not to use genAI) may have accessed such tools, potentially influencing the results.

6 Conclusion

This research affirms the importance of balancing genAI assistance with foundational pedagogical skills, ensuring that novice instructional designers critically evaluate AI-generated content for alignment with best practices and contextual relevance. By having these novice instructional designers develop microlessons both with and without genAI, we demonstrated that its thoughtful integration can yield higher-quality instructional materials. As the instructional design process continues to evolve, educators and designers must remain mindful of how genAI is deployed during coursework, ensuring it enhances rather than replaces fundamental learning experiences. Our findings highlight the potential of genAI to support a future in which genAI complements human expertise, working together to create more effective, engaging, and accessible learning experiences.

Acknowledgments. We thank the Eberly Center for Teaching Excellence and Educational Innovation for their support via the Generative Artificial Intelligence Teaching as Research initiative.

References

1. Bier, N., Stamper, J., Moore, S., Siegel, D., Anbar, A.: OLI Torus: a next-generation, open platform for adaptive courseware development, delivery, and research. In: Companion Proceedings of the 13th International Learning Analytics and Knowledge Conference. (2023).
2. Buchem, I., Hamelmann, H.: Microlearning: a strategy for ongoing professional development. ELearning Pap. 21, 7, 1–15 (2010).
3. Chen, X., Wu, D.: Automatic generation of multimedia teaching materials based on generative AI: Taking Tang poetry as an example. IEEE Trans. Learn. Technol. (2024).
4. Debattista, M.: A comprehensive rubric for instructional design in e-learning. Int. J. Inf. Learn. Technol. 35, 2, 93–104 (2018).
5. Denny, P., Gulwani, S., Heffernan, N.T., Käser, T., Moore, S., Rafferty, A.N., Singla, A.: Generative AI for Education (GAIED): Advances, Opportunities, and Challenges, http://arxiv.org/abs/2402.01580, (2024). https://doi.org/10.48550/arXiv.2402.01580.
6. Fleiss, J.L., Cohen, J.: The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. Educ. Psychol. Meas. 33, 3, 613–619 (1973). https://doi.org/10.1177/001316447303300309.
7. Guo, Z., Liu, Z., Ge, S., Li, Y.: Effectiveness of Contextualized Micro-lessons in the Acquisition of College English Vocabulary. Int. J. Engl. Lang. Educ. Lit. Stud. 2, 4, (2023).
8. Handayani, R.D.: Online microteaching lesson study: a recipe to enhance prospective physics teachers' pedagogical knowledge. Int. J. Lesson Learn. Stud. 11, 3, 221–234 (2022).
9. Hug, T.: Mobile learning as' microlearning': Conceptual considerations towards enhancements of didactic thinking. Int. J. Mob. Blended Learn. IJMBL. 2, 4, 47–57 (2010).
10. Huong, P.T., Linh, C.T.H.: Building Rubrics for Evaluating the Competence of Preparing for Lesson Plans of Pedagogical Student. World. 5, 5, 175–179 (2017).
11. Ilic, P.: Micro-lessons as a response to emergency remote teaching. In: 2022 IEEE Learning with MOOCS (LWMOOCS). pp. 155–160 IEEE (2022).
12. Ilic, P.: Optimizing Education Through Micro-Lessons: Engaging and Adaptive Learning Strategies: Engaging and Adaptive Learning Strategies. IGI Global (2024).
13. Janssen, F.J., Westbroek, H.B., van Driel, J.H.: How to make guided discovery learning practical for student teachers. Instr. Sci. 42, 67–90 (2014).
14. Kim, J.: Learnersourcing: improving learning with collective learner activity. Massachusetts Institute of Technology (2015).
15. Kohavi, R., Longbotham, R.: Online controlled experiments and A/B tests. Encycl. Mach. Learn. Data Min. 1–11 (2015).
16. Kumar, S., Gunn, A., Rose, R., Pollard, R., Johnson, M., Ritzhaupt, A.D.: The Role of Instructional Designers in the Integration of Generative Artificial Intelligence in Online and Blended Learning in Higher Education. Online Learn. 28, 3, 207–231 (2024).
17. Laal, M., Laal, M.: Collaborative learning: what is it? Procedia-Soc. Behav. Sci. 31, 491–495 (2012).
18. Lim, C.P., Chan, B.C.: MicroLESSONS in teacher education: Examining pre-service teachers' pedagogical beliefs. Comput. Educ. 48, 3, 474–494 (2007).
19. Monib, W.K., Qazi, A., Apong, R.A.: Mapping microlearning development and trends across diverse contexts: a bibliometric analysis (2007–2023). Interact. Learn. Environ. 1–46 (2024). https://doi.org/10.1080/10494820.2024.2402556.
20. Moore, S., Bier, N., Stamper, J.: Assessing Educational Quality: Comparative Analysis of Crowdsourced, Expert, and AI-Driven Rubric Applications. In: Proceedings of the AAAI Conference on Human Computation and Crowdsourcing. pp. 115–125 (2024).

21. Moore, S., Nguyen, H.A., Stamper, J.: Examining the Effects of Student Participation and Performance on the Quality of Learnersourcing Multiple-Choice Questions. In: Proceedings of the Eighth ACM Conference on Learning @ Scale. pp. 209–220 ACM, Virtual Event Germany (2021). https://doi.org/10.1145/3430895.3460140.
22. Nikou, S.A., Economides, A.A.: Mobile-Based micro-Learning and Assessment: Impact on learning performance and motivation of high school students. J. Comput. Assist. Learn. 34, 3, 269–278 (2018). https://doi.org/10.1111/jcal.12240.
23. Palmon, I., Brown, C.S., Highet, A., Kulick, A.A., Barrett, M.E., Cassidy, D.E., Herman, A.E., Gomez-Rexrode, A.E., O'Reggio, R., Sonnenday, C.: Microlearning and social media: a novel approach to video-based learning and surgical education. J. Grad. Med. Educ. 13, 3, 323–326 (2021).
24. Parsons, B., Curry, J.H.: Can ChatGPT pass graduate-level instructional design assignments? Potential implications of artificial intelligence in education and a call to action. TechTrends. 68, 1, 67–78 (2024).
25. Rose, D.: Universal Design for Learning. J. Spec. Educ. Technol. 15, 4, 47–51 (2000). https://doi.org/10.1177/016264340001500407.
26. Singh, A., Brooks, C., Lin, Y., Li, W.: What's In It for the Learners? Evidence from a Randomized Field Experiment on Learnersourcing Questions in a MOOC. In: Proceedings of the Eighth ACM Conference on Learning @ Scale. pp. 221–233 ACM, Virtual Event Germany (2021). https://doi.org/10.1145/3430895.3460142.
27. Stahl, E., Bromme, R.: Not everybody needs help to seek help: surprising effects of metacognitive instructions to foster help-seeking in an online-learning environment. Comput. Educ. 53, 4, 1020–1028 (2009).
28. Torres, K.M.: Tiny Lessons, Massive Gains: A New Approach to Learning. In: Optimizing Education Through Micro-Lessons: Engaging and Adaptive Learning Strategies. pp. 325–341 IGI Global (2024).
29. Torres, K.M.: Transforming Higher Education With Microlessons. In: Global Perspectives on Micro-Learning and Micro-Credentials in Higher Education. pp. 59–74 IGI Global (2024).