Integrating Generative AI into Instructional Design Practice: Effects on Graduate Student Learning and Self-Efficacy

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
[2] Eberly Center Teaching Excellence & Educational Innovation, Pittsburgh PA 15213, USA

Abstract. Generative AI (genAI) tools are increasingly being integrated into instructional design workflows for content creation, assessment development, and lesson planning. For novice designers, it is critical to understand whether this integration supports the design process without compromising underlying pedagogical learning. This study addresses this gap through an ecologically valid field experiment within a 14-week graduate course training novice instructional designers. Using a counterbalanced A/B design embedded in authentic coursework, students created eight microlessons, alternating genAI assistance with independent work. Learning of pedagogical principles was assessed via module pre/post-tests and self-efficacy, in teaching practices and genAI use, was measured via course-level pre/post-surveys. Results showed no evidence that genAI use hindered learning, as post-test scores on module content remained stable or improved, despite variations in test form difficulty. Students demonstrated substantial and statistically significant gains by over 10% in self-efficacy related to both their teaching practices and their ability to leverage genAI. By evaluating genAI within sustained and authentic instructional design activities, this study demonstrates that thoughtfully integrated AI tools can effectively support novice designers, enhancing their confidence and professional growth without compromising foundational pedagogical learning.

Keywords: Instructional Design, Design for Learning, Generative AI, Self-Efficacy, A/B Study.

1 Introduction

Generative AI (genAI) tools are transforming educational environments, influencing both student work quality and the learning process itself [9, 10]. While offering benefits like idea generation and real-time feedback, genAI may shift cognitive demands of academic tasks [24]. This is particularly salient in project-based learning environments, where students are expected to apply pedagogical principles to design educational materials. In such contexts, genAI can increase efficiency and streamline production, but may simultaneously reduce opportunities for deep engagement and conceptual understanding [37]. At the same time, it may shape students' confidence in their capabilities by enhancing self-efficacy when used successfully or diminishing it if students become overly reliant on AI-generated content [38]. As educators and institutions begin integrating genAI into the classroom, it is essential to understand its implications not only for productivity, but for learner growth.

However, the use of genAI in education may also present challenges, particularly by limiting creativity and, for students, impeding deeper learning as they engage superficially with content [39]. Over-reliance on AI can hinder the development of foundational teaching and learning skills, overshadowing the growth of critical thinking and creativity. As more practitioners and educators integrate genAI into assignments to enhance student engagement, students themselves are increasingly using these tools, sometimes even without explicit instruction to do so [19]. This growing reliance on AI may influence how students approach assignments, including microlesson creation, and could impact their confidence in their own abilities as well as their proficiency with these tools as educators.

This study examines genAI's impact on student learning and self-efficacy within a graduate course where students created eight microlessons [35] incorporating distinct learning science principles through an A/B experimental design. For each microlesson, students alternated between using genAI assistance and working independently when generating the learning objectives, assessments, and instructional content. Student learning gains were assessed through pre- and post-tests for each set of microlessons. Additionally, students' self-efficacy as educators and as users of genAI was measured through pre- and post-course assessments. We investigated the following two research questions:

RQ1: In what ways does employing genAI during the creation of microlessons affect students' learning of fundamental teaching and learning principles?

RQ2: How do graduate students' self-efficacy as educators and users of genAI change over the duration of a graduate-level course that integrates genAI tasks?

Through the investigation of these research questions, this work makes the following contributions: First, we show that employing genAI does not compromise students' learning of teaching principles. Second, we provide evidence that students' self-efficacy in both teaching practices and genAI usage increases significantly when they engage with genAI-assisted instructional design. Third, we present a novel, ecologically valid implementation of genAI within a graduate-level course on instructional design. This research offers insights into how future instructional designers learn and evolve with AI tools in authentic, real-world contexts.

2 Related Work

2.1 Learning and Self-Efficacy

Engaging students in the creation of microlessons is not just a practical skill-building exercise, it is also a powerful strategy for learning [13]. Design-based pedagogies, such as project- and problem-based learning require students to apply theoretical knowledge to authentic tasks, promoting deeper cognitive engagement and knowledge transfer [17, 21]. When students are asked to create instructional content, they must analyze subject matter, structure ideas coherently, and consider how others will interact with the material. These processes enhance metacognitive awareness and encourage students to transform information into usable knowledge [2]. Beyond cognitive benefits, design-based learning tasks also support the development of self-efficacy, students' beliefs in their ability to perform specific tasks successfully. Social cognitive theory identifies mastery experiences as the most influential source of self-efficacy, and instructional design activities often provide precisely these experiences [32]. By working through iterative cycles of creation, revision, and feedback, students gain confidence in both their subject-matter understanding and their professional identity as educators or designers [28, 40].

However, the integration of genAI into the design process complicates these dynamics. While AI-enhanced tools can support students by reducing cognitive load or enhancing surface-level quality, they may also alter the conditions under which self-efficacy develops. For instance, if students attribute successful performance to the genAI rather than their own effort or understanding, their self-efficacy may stagnate or even decline [26]. Conversely, students who learn to use genAI critically, by evaluating and refining outputs rather than passively accepting them, may gain confidence in both their instructional judgment and their capacity to harness emerging technologies effectively [25]. Understanding genAI's impact on learning and self-efficacy in design-based educational settings is essential, as student interpretation of progress within AI-enhanced environments significantly affects learning outcomes [36].

2.2 GenAI's Influence on Instructional Design

GenAI is transforming instructional design by enhancing creativity, streamlining planning, and providing personalized feedback while catalyzing idea generation for diverse learning needs [8, 11]. AI tools expedite design processes through innovative suggestions for lessons, assessments, and materials while improving quality through efficient iteration and refinement [22, 29]. Recent studies have explored genAI applications in instructional design, such as using tools like ChatGPT to create course materials [7, 15]. For example, Choi et al. [8] found that experienced instructional designers using ChatGPT for course mapping identified it as providing solid baselines for creating effective and efficient course structures, though still requiring domain expertise for accuracy and contextual relevance. Another investigation assessed ChatGPT's potential in developing lesson plans, showing that the tool could offer valuable starting points by generating general structures, ideas, and resources, which is an especially useful feature for novice educators [3]. Since teachers must evaluate the AI-generated content to ensure its alignment with learning objectives, researchers suggested that engaging with ChatGPT might also enhance teachers' critical thinking skills in lesson planning [14].

2.3 Evolving Roles in Instructional Design

The role of instructional designers is undergoing a significant transformation with the rise of genAI. Krushinskaia et al. [22] proposed a study to explore the potential of genAI as a co-designer in instructional settings, focusing on how teachers could collaborate with LLMs fine-tuned for educational purposes, such as Google's LearnLM [34]. Their research aims to investigate whether such partnerships could provide students with more personalized and effective learning opportunities. Despite consensus on ChatGPT's utility for instructional design [5, 12], designers must critically assess AI outputs using foundational knowledge to ensure pedagogical alignment [3]. This reliance highlights a key challenge that successful collaboration with AI requires a strong grasp of instructional design principles, yet research shows that even experienced educators often lack robust instructional design proficiency. This gap becomes even more pronounced for novice educators and learning engineers, who may struggle to effectively evaluate and refine AI-generated content [16].

3 Methods

The study was conducted as part of a 14-week graduate-level course on Tools for Online Learning during the spring 2024 semester at a large university in the northeastern United States. This course covers the topics of educational technology and learning science principles. The course met twice a week for 80 minute in-person class sessions, which typically consisted of lectures supported by a slide deck, group discussion, and active learning activities. Outside of class, students completed biweekly projects providing hands-on experience with different pieces of educational technology and completed online course materials, consisting of instructional content and low-stakes assessments, before class time. Our intervention in this course spanned four week-long modules covering the following topics: Universal Design for Learning (UDL) [27], Guided Discovery (GD) [18], Fostering Help-Seeking (FHS) [31], and Collaborative Learning (CL) [23].

3.1 Participants

Participants were 27 second-semester master's students (ages 22-36) enrolled in a master's level Educational Technology and Applied Learning Sciences program. All students had foundational knowledge from first-semester coursework and prior genAI experience, representing emerging instructional design practitioners.

3.2 Learning Platform

The four modules from the course used in this study were part of the online materials students completed before class time, delivered via the Open Learning Initiative (OLI) platform, a well-established open-ended learning environment that offers courses across various domains, including anatomy, foreign language, and statistics [4]. The study modules were delivered via the Open Learning Initiative (OLI) platform, which integrates instructional content with both low-stakes formative and high-stakes summative assessments. Each module in the course was designed to be completed over approximately one week and was structured similarly to textbook chapters, containing three to six related topics. The instructional content was presented across multiple pages, including text and brief instructional videos. Interspersed throughout were formative activities, such as multiple-choice and short-answer questions, which provided students with opportunities for practice and feedback. Students could attempt these optional activities multiple times, receiving instant feedback with each attempt. Each module contained prompts for students to create and submit two microlessons for a total of eight microlessons overall.

3.3 Design

This study investigated the impact of integrating genAI into a core instructional design activity embedded within an authentic graduate course. We utilized the required coursework task of creating microlessons as the specific context to explore how genAI assistance influences student learning of pedagogical principles and their self-efficacy regarding both instructional design and genAI use. The study followed an A/B counterbalanced experimental design [20], where students created eight microlessons across four course modules. For each module's pair of microlessons, students were randomly assigned to use ChatGPT (built upon GPT-4) as an assistant for one microlesson (treatment condition) and to complete the other without AI assistance (control condition). The condition assignment alternated across modules in a counterbalanced manner, ensuring that by the end of the course, each student had completed four microlessons with AI support and four without. This within-subjects design allowed for direct comparison of outcomes associated with using genAI during the instructional design process versus traditional methods.

The creation of each microlesson served as a practical exercise in lesson planning and construction, following the fundamental instructional design principle of backwards design. Students developed their microlessons within the course's learning management system, Canvas, following a four-step scaffolded process. Each step was associated with an essay question that allowed students to enter and format text. At the outset, students selected a topic for their microlesson and were reminded of the specific learning science principle they were required to incorporate. In the next step, they formulated one or more learning objectives, ensuring that these objectives were actionable, measurable, and student-centered. In the third step, students created at least three assessments in any format they preferred, such as multiple-choice or short-answer questions. Finally, they wrote one to five paragraphs of instructional content aligned with their chosen topic and learning objectives, integrating a strategy to apply the designated learning science principle.

Each microlesson was required to incorporate a specific learning science principle covered in the corresponding module, in either the instructional paragraphs or through the three assessment items. For example, in the first module on UDL, the two microlessons students created needed to incorporate one of the seven principles or one of the eight steps of UDL [6, 33]. A list of the eight learning science principles incorporated into the different microlessons shown in order can be seen in Table 1.

Table 1. Overview of the four modules paired with the eight learning science principles used for micro-lesson development.

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

3.4 Generative AI Usage by Students

For microlessons completed under the genAI condition, students were explicitly guided to use ChatGPT to support their design process, simulating realistic human-AI collaboration in instructional design. Instructions accompanying each genAI-assigned microlesson prompted students to leverage ChatGPT for specific tasks across the design process: (1) constructing actionable, measurable learning objectives, (2) brainstorming applications of the designated learning science principle within assessments or instruction, (3) creating or refining formative assessment questions related to the objectives, and (4) drafting or editing instructional text, focusing on clarity and appropriate tone. To document this collaboration, students were required to submit a transcript of their ChatGPT conversation via shareable link or text document. While a detailed analysis of these interaction logs is beyond the scope of this paper, Table 2 provides illustrative examples of common student prompting strategies observed during this process.

Table 2. Examples of prompts used by students and their intended purpose.

Purpose	Student Prompt Examples
Brainstorming learning objectives	What might be a few learning objectives for a lesson on conjugating verbs in Spanish?
Simplifying complex concept	Explain the testing effect and stereotype threat to someone with no background in psychology
Creating analogies or examples	Please give me a real world example for the concept of

	interleaved practice
Drafting lesson text	Write an introduction to a 5-minute lesson on indexing into a list in python programming
Refining quiz questions	Write three multiple choice questions with distractors for a lesson on the effects of marine life on…
Revising explanations	Make this explanation clearer: 'Learners remember better when they...'
Prompt chaining for quality	That example was too confusing can you give a more concrete one?

To maintain a focus on pedagogical intentionality, students were continually reminded to critically evaluate all AI-generated outputs for accuracy, pedagogical soundness, and alignment with learning objectives. The instructions emphasized that genAI should serve as a tool to assist their instructional design decisions, not replace their critical judgment. To maintain ecological validity and capture authentic usage patterns within this real-course setting, we deliberately allowed flexibility, as no constraints were imposed on the number of student-AI interactions, nor were specific prompt structures mandated.

Additionally, for microlessons assigned to the non-genAI (control) condition, students received explicit instructions not to use any genAI tools. To encourage adherence in this setting, we implemented several measures to help keep the students honest. Students confirmed compliance via a checkbox upon submission and were reassured that assignments would not be penalized for lacking surface polish, such as the use of informal language or the presence of minor typos, in attempts to reduce potential incentives to use AI purely for refinement. Instructors also qualitatively reviewed submissions noting typical stylistic differences, such as simpler sentence structures and less elaborate phrasing in the non-genAI work, which provided some indication of adherence. Despite these measures, we acknowledge the inherent limitation that total adherence to the non-genAI condition cannot be guaranteed outside a controlled laboratory environment, which is considered in the interpretation of our findings.

3.5 Data Sources

Self-Efficacy. The course also incorporated an assessment of students' self-efficacy in both using AI tools and applying learning science principles in teaching both at the beginning and the end of the course. Students were presented with six statements (3 measuring AI self-efficacy, 3 measuring self-efficacy on their ability to design educational lessons using learning principles) and asked to indicate their level of confidence in their abilities on a 100-point scale (in increments of 10) with higher numbers expressing a higher degree of self-perceived confidence. The same set of questions was used in both the pre-test and post-test, allowing for a direct comparison over time.

Knowledge Checks. Additionally, at the beginning of each module, students completed a pre-test consisting of five multiple-choice questions intended to assess students' knowledge of the module's key concepts. Upon completing the module, they took a post-test with a different set of five questions targeting the same concepts (i.e., these items were designed to be isomorphic). For these pre- and post-tests, an A/B design was implemented, where half the students received version A of the pre-test and version B of the post-test, while the other half received version B of the pre-test and version A of the post-test. These four sets of pre/post-tests served as a measure of learning gains as students progressed through the given modules. Two questions on each module's pre- and post-test specifically assessed understanding of the learning science principles applied in that module's microlessons. Since students used genAI for one of the microlessons per module, these targeted questions enabled comparing learning outcomes between the AI-assisted and non-AI conditions for each principle.

4 Results

4.1 Student Learning with GenAI

Student learning gains were evaluated using a 2 (pre–post) × 2 (test form sequencing: A→B vs. B→A) mixed-model ANOVA. In this design, half the students received test form A at pre-test and form B at post-test, while the remaining half experienced the reverse order. Assignment to test order was independent of the genAI conditions. The average scores by test form sequence across all four modules can be seen in Figure 1.

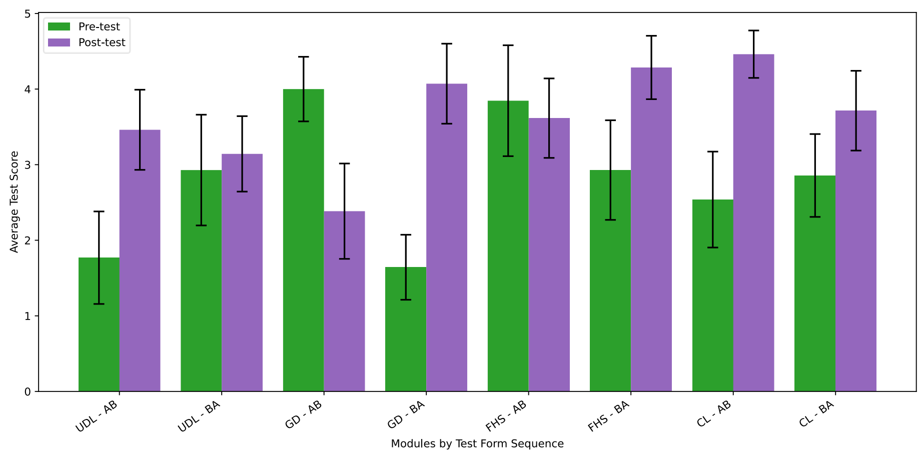

Fig. 1. Pre- and post-test average total scores across all four modules, segmented by test form sequence (AB or BA).

Analyses revealed significant interactions between time and test sequence across modules, complicating interpretations due to likely test form difficulty differences (see Table 3). For UDL, overall improvement was significant but driven only by the A→B group, with pre-test scores suggesting Form A was harder. For GD, no overall time effect was found, but an interaction showed the B→A group improved while A→B declined, with pre-tests suggesting Form A was easier. Similarly, FHS showed overall improvement, but the interaction indicated only the B→A group gained significantly (again, pre-tests suggested Form A was easier). Conversely, the CL module showed strong overall gains, with an interaction favoring the A→B group, leading to significant post-test differences. These complex interactions consistently highlight test form difficulty as a likely confound in observed learning gains.

Table 3. ANOVA results for quiz scores by module, * indicates significance at $p < .05$.

Module	Time Effect	Test Order Effect	Interaction
Universal Design for Learning (UDL)	12.47, p = .002* (η^2_p = .33)	2.16, p = .15 (η^2p = .08)	7.50, p = .01* (η^2p = .23)
Guided Discovery (GD)	1.89, p = .18 (η^2p = .33)	3.35, p = .08 (marginal; η^2p = .12)	59.38, p < .001* (η^2p = .70)
Fostering Help-Seeking (FHS)	4.70, p = .04* (η^2p = .16)	0.19, p = .67 (η^2p = .007)	9.33, p = .005* (η^2p = .27)
Collaborative Learning (CL)	37.61, p < .001* (η^2p = .60)	0.71, p = .41 (η^2p = .03)	5.53, p = .03* (η^2p = .18)

Targeted MCQs to assess impact of genAI on knowledge of learning science principles. To assess the impact of genAI assistance on knowledge of specific learning science principles targeted in each module, we analyzed pre-post change scores (improvement, no change, or decline) on two specific multiple-choice questions included in the pre- and post-tests for each module. These questions directly assessed understanding of the principle students were required to implement in their corresponding microlessons for that module. We conducted Mann-Whitney U-tests comparing the change scores on each targeted item between the group that used genAI for the associated microlesson and the group that did not. The results of these comparisons are presented in Table 4.

Table 4. A comparison of pre/post-test change scores on targeted module questions using Mann-Whitney U-Tests, where a * denotes a significant difference.

Module Topic	Principle	U Value	*p*-Value
Universal Design for Learning (UDL)	Seven Principles	123	0.13
	Eight Steps	74	0.43
Guided Discovery (GD)	Contrasting Cases	124.5	0.10
	Tell-then-Practice	136.5	0.03*
Fostering Help-Seeking (FHS)	Strategies to Improve	96.5	0.79
	Leveraging Prior Knowledge	153	0.002*
Collaborative Learning (CL)	Cooperative Learning	74.5	0.43
	Peer Learning	42	0.02*

As shown in Table 4, the results examining the impact of genAI use on learning specific principles were mixed. Significant differences favoring the genAI group emerged for the second microlesson items in the FHS (p = .002) and CL (p = .02) modules, with a marginal difference favoring the AI group for the first GD item (p = .10). However, for the second GD item, the non-genAI group showed significantly greater improvement (p = .03). Several comparisons yielded no significant differences between the groups. While these findings point to some potential benefits of using genAI during the instructional design process for learning the targeted principles, we remain cautious in their interpretation. In each case where a significant difference was found, the group that performed better at post-test also happened to have the easier test form for that specific item, making it difficult to attribute improvements solely to genAI use. Nonetheless, these trends suggest this is a promising area for further investigation, ideally employing more standardized assessment measures to isolate the effect of genAI on learning core pedagogical concepts.

4.2 Students Self-Efficacy

We assessed student self-efficacy using a six-item survey administered at the start (pre-test) and end (post-test) of the course. The survey comprised two three-item subscales: one focused on teaching practices and the other on the use of genAI. Internal consistency reliability for both subscales was evaluated using McDonald's omega (ω) and found to be acceptable at both pre- and post-test. To evaluate changes in self-efficacy over the semester, paired t-tests were conducted comparing pre-test and post-test scores for each subscale. Table 5 presents the descriptive statistics, reliability coefficients, and the results of these paired t-test analyses across the 27 students.

Table 5. Descriptive statistics, reliability, and pre/post-test changes in self-efficacy measures.

Self-Efficacy Measure	Pre-Test Mean (SD)	Post-Test Mean (SD)	Pre-Test ω	Post-Test ω	Pre/Post Change	Effect Size (Hedges' g)
Teaching Practices	68.02 (14.18)	78.52 (11.67)	.84	.74	$t(26) = -5.53$, $p < .001$	1.03
genAI Use	69.75 (18.58)	81.72 (12.38)	.91	.71	t(26) = -5.03, $p < .001$	0.94

Students reported statistically significant increases in self-efficacy from pre-test to post-test for both teaching practices and the use of genAI, with large effect sizes observed for both changes (Hedges' g = 1.03 and 0.94, respectively). On average, scores for both measures increased by more than 10 points over the semester. Additionally, we compared the two types of self-efficacy within participants at each time point using paired t-tests. There was no significant difference between teaching practice self-efficacy and genAI use self-efficacy at the start of the course ($t(26) = -0.89$, p = .38, Hedges' g = -0.17), nor was there a significant difference between them at the end of the course ($t(26) = -1.37$, p = .18, Hedges' g = -0.26). Students showed substantial, comparable self-efficacy growth in both domains throughout the semester.

5 Discussion

This study investigated how novice instructional designers and learning engineers used genAI to design microlessons within an authentic graduate-level course setting. Across four week-long modules, students collaborated with or without genAI to create instructional materials, while engaging in a parallel learning experience about core learning science principles. Our findings indicate that students generally improved on quizzes assessing fundamental learning and learning science principles from pre to post across the four modules. However, variations in test form difficulty warrant caution in drawing definitive conclusions about content mastery. Additionally, students reported significant gains in self-efficacy from the beginning to the end of the course, both in their teaching practices and their ability to integrate genAI into the lesson design workflow. These results suggest that structured integration of genAI into instructional design coursework can occur without hindering student learning of core principles, while simultaneously enhancing confidence in both pedagogical skills and AI tool utilization.

A key contribution of this study lies in its integration of genAI into an authentic semester-long graduate course setting. This environment moves beyond short-term or simulated use cases to examine how genAI is integrated into the actual workflow of novice instructional designers dealing with real course requirements, deadlines, and iterative feedback. This provides ecologically valid insights into how emerging instructional designers engage with genAI tools over time. We simultaneously evaluated the impact of genAI assistance on (a) learning outcomes related to core pedagogical principles using targeted pre-post assessments, (b) student self-efficacy concerning both instructional design practices and genAI use, and (c) the instructional design process itself via a counter-balanced A/B design comparing AI-assisted vs. non-AI work on the same types of tasks. The work provides a holistic picture of genAI's role during instructional design, going beyond a single dimension such as output quality or user perception. It provides specific evidence regarding how future professionals in education technology and learning sciences develop skills and confidence when learning with and about genAI tools integrated into their core training.

Interpreting these results requires considering factors like prior knowledge, engagement, experience, and particularly confounding variations in test form difficulty. While this prevents definitive causal claims about genAI enhancing knowledge based solely on quiz scores, the stable or improved post-test performance, combined with increased self-efficacy, suggests genAI support did not impede learning. A key finding is that the structured use of generative AI did not hinder students' learning of core pedagogical principles. Although post-test scores were stable or improved, our analysis did not show consistent, statistically significant learning gains attributable to the genAI condition, a result confounded by variations in test form difficulty. It is possible that the primary benefit of the AI intervention was allowing students to focus on higher-order tasks like structuring and refinement, an advantage not fully captured by our knowledge-based assessments. This suggests that while genAI can be safely integrated, its main contribution may be in enhancing the design workflow rather than directly accelerating conceptual understanding. This finding aligns with research showing that creating educational content fosters understanding [1, 30]. The use of genAI might facilitate this by allowing students to focus more on higher-order tasks like structuring and refinement. Nonetheless, effective instructional design still hinges on pedagogical expertise for critically evaluating and adapting AI-generated materials to ensure quality and contextual fit [29]. Our findings, while positive regarding learning outcomes, do not eliminate concerns that over-reliance on genAI could hinder the development of foundational pedagogical skills. Although students' improved post-test performance and increased self-efficacy are promising, this work highlights a potential gap in skill development. Facilitating this growth may require strategies like thoughtful reflection and critical evaluation of AI-generated materials.

The self-efficacy pre-post tests revealed significant growth in students' confidence both as educators and as users of genAI. Although the study lacks a control group entirely unexposed to genAI, as all students participated in the course rotation, this consistent increase across two distinct self-efficacy measures provides strong indicative evidence of positive development. This is particularly relevant as tools like ChatGPT become commonplace in professional lesson planning [3]. While such tools offer efficiency in instructional design, concerns remain about whether novice designers can maintain quality control and retain agency over AI-generated content. However, our results suggest that integrating genAI into microlesson development strengthened students' confidence in both pedagogical strategy and AI use. This gain was reflected in the student self-efficacy reports, supporting the idea that well-designed AI integration can be a catalyst for professional growth among instructional designers.

6 Limitations and Future work

In our research, we conducted an A/B field study in a real class for novice instructional designers and learning engineers. However, we acknowledge this work was conducted in a single course with a relatively small sample size, which may restrict the generalizability of our findings. The ecological validity inherent in this field study design presents challenges in ensuring perfect experimental control. Specifically, we must acknowledge the possibility of control condition leakage, where students instructed not to use genAI might have done so despite monitoring efforts (detailed previously in the Methods section). In our experimental design, we deliberately used different test forms at pre- and post-test to reduce practice effects and assess generalizable learning gains. Unfortunately, these test forms ended up differing in difficulty, leading to a potential ceiling effect in pre-test scores that complicates the interpretation of content mastery in certain modules. Additionally, while self-efficacy improvements were observed, the absence of a control group for this measure limited our ability to draw firm conclusions about the sustained impact of genAI on confidence with instructional design skills relative to a group with no AI exposure at all. Although our counter-balanced, within-subjects design let every student serve as their own control, every participant ultimately engaged with genAI for half of the microlessons. This raises concerns regarding spill-over, where strategies or content created with genAI may have influenced work in the nominal no-genAI condition, and course-level motivational effects that cannot be disentangled from genAI exposure.

In future research, we aim to address these limitations by incorporating a larger, more diverse sample across multiple courses and educational settings. Longitudinal studies tracking both educational artifact quality and self-efficacy over time can help determine whether improvements continue as students gain more experience with genAI. We plan to investigate how specific rubric criteria differ between genAI-assisted and non-assisted designs, such as the quality of learning objectives, to provide more nuanced insights into the effects of genAI integration. These efforts will help clarify the pedagogical benefits and potential trade-offs of incorporating genAI into the instructional design process.

7 Conclusion

This study investigated the integration of genAI into authentic instructional design tasks within a graduate course for novice instructional designers. Our findings demonstrate that structured use of genAI for creating microlessons did not impede students' learning of core pedagogical principles, while significantly boosting their self-efficacy in both general teaching practices and the specific use of AI tools for instructional design. Our approach embedded AI use within real coursework over a semester and employed a comparative approach to assess learning alongside self-efficacy. This yielded ecologically valid insights regarding the ways emerging professionals engage with AI tools and develop confidence through practice. These results suggest that thoughtful genAI integration can support novice instructional designers' development, enhancing their confidence and potentially streamlining design processes without sacrificing foundational knowledge. This work demonstrates the potential for genAI to function as a valuable collaborator in instructional design, fostering growth when implemented within authentic, pedagogically sound learning experiences that balance technological assistance with human expertise.

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

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Aflalo, E.: Students generating questions as a way of learning. Act. Learn. High. Educ. 22, 1, 63–75 (2021).
2. Barron, B., Darling-Hammond, L.: Teaching for meaningful learning: A review of research on inquiry-based and cooperative learning. Book excerpt. Educ. Found. (2008).

3. van den Berg, G., du Plessis, E.: ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education. Educ. Sci. 13, 10, 998 (2023).
4. Bier, N. et al.: 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).
5. Bolick, A.D., Da Silva, R.L.: Exploring artificial intelligence tools and their potential impact to instructional design workflows. TechTrends. 68, 1, 91–100 (2024).
6. Burgstahler, S.: Universal Design: Process, Principles, and Applications. -IT. (2009).
7. Chen, W. et al.: A survey study of kindergarten teachers' use of ChatGPT to support instructional design. J. Educ. Technol. Innov. 6, 3, (2024).
8. Choi, G.W. et al.: Utilizing Generative AI for Instructional Design: Exploring Strengths, Weaknesses, Opportunities, and Threats. TechTrends. 68, 4, 832–844 (2024).
9. Cipriano, B.P. et al.: A Picture Is Worth a Thousand Words: Exploring Diagram and Video-Based OOP Exercises to Counter LLM Over-Reliance. In: Ferreira Mello, R. et al. (eds.) Technology Enhanced Learning for Inclusive and Equitable Quality Education. pp. 75–89 Springer Nature Switzerland, Cham (2024).
10. Denny, P. et al.: Generative AI for Education (GAIED): Advances, Opportunities, and Challenges, (2024). https://doi.org/10.48550/arXiv.2402.01580.
11. Farrokhnia, M. et al.: A SWOT analysis of ChatGPT: Implications for educational practice and research. Innov. Educ. Teach. Int. 61, 3, 460–474 (2024). https://doi.org/10.1080/14703297.2023.2195846.
12. Grassini, S.: Shaping the future of education: exploring the potential and consequences of AI and ChatGPT in educational settings. Educ. Sci. 13, 7, 692 (2023).
13. Handayani, R.D.: Online microteaching lesson study: a recipe to enhance prospective physics teachers' pedagogical knowledge. Int. J. Lesson Learn. Stud. 221–234 (2022).
14. Hodges, C.B., Kirschner, P.A.: Innovation of Instructional Design and Assessment in the Age of Generative Artificial Intelligence. TechTrends. 68, 1, 195–199 (2024). https://doi.org/10.1007/s11528-023-00926-x.
15. Hu, B. et al.: Teaching Plan Generation and Evaluation With GPT-4: Unleashing the Potential of LLM in Instructional Design. IEEE Trans. Learn. Technol. (2024).
16. Huizinga, T. et al.: Teacher involvement in curriculum design: need for support to enhance teachers' design expertise. J. Curric. Stud. 46, 1, 33–57 (2014). https://doi.org/10.1080/00220272.2013.834077.
17. Hung, W. et al.: Problem-based learning. In: Handbook of research on educational communications and technology. pp. 485–506 Routledge (2008).
18. Janssen, F.J. et al.: How to make guided discovery learning practical for student teachers. Instr. Sci. 42, 67–90 (2014).
19. Kazemitabaar, M. et al.: CodeAid: Evaluating a Classroom Deployment of an LLM-based Programming Assistant that Balances Student and Educator Needs. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. pp. 1–20 ACM, Honolulu HI USA (2024). https://doi.org/10.1145/3613904.3642773.
20. Kohavi, R., Longbotham, R.: Online controlled experiments and A/B tests. Encycl. Mach. Learn. Data Min. 1–11 (2015).
21. Kolodner, J.L. et al.: Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting learning by design (tm) into practice. J. Learn. Sci. 12, 4, 495–547 (2003).
22. Krushinskaia, K. et al.: Effects of Generative Artificial Intelligence on Instructional Design Outcomes and the Mediating Role of Pre-service Teachers' Prior Knowledge of Dif-

ferent Types of Instructional Design Tasks. In: Olney, A.M. et al. (eds.) Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky. pp. 395–400 Springer Nature Switzerland, (2024). https://doi.org/10.1007/978-3-031-64312-5_49.
23. Laal, M., Laal, M.: Collaborative learning: what is it? Procedia-Soc. Behav. Sci. 31, 491–495 (2012).
24. Nazaretsky, T. et al.: AI or Human? Evaluating Student Feedback Perceptions in Higher Education. In: Ferreira Mello, R. et al. (eds.) Technology Enhanced Learning for Inclusive and Equitable Quality Education. pp. 284–298 Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-72315-5_20.
25. Power, R.: Evaluating Graduate Education Students' Self-Efficacy with the Use of Artificial Intelligence Agents: A Case Study. J. Educ. Inform. 5, 1, 3–19 (2024).
26. Rodríguez-Ruiz, J. et al.: Is artificial intelligence use related to self-control, self-esteem and self-efficacy among university students? Educ. Inf. Technol. 30, 2, 2507–2524 (2025).
27. Rose, D.: Universal Design for Learning. J. Spec. Educ. Technol. 15, 4, 47–51 (2000). https://doi.org/10.1177/016264340001500407.
28. Roytek, M.A.: Enhancing instructional design efficiency: Methodologies employed by instructional designers. Br. J. Educ. Technol. 41, 2, 170–180 (2010).
29. Ruiz-Rojas, L.I. et al.: Empowering education with generative artificial intelligence tools: Approach with an instructional design matrix. Sustainability. 15, 15, 11524 (2023).
30. Singh, A. et al.: 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.
31. 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).
32. Stajkovic, A.D., Luthans, F.: Social cognitive theory and self-efficacy: Goin beyond traditional motivational and behavioral approaches. Organ. Dyn. 26, 4, 62–74 (1998).
33. Story, M.F.: Principles of universal design. Univers. Des. Handb. 2, (2001).
34. Team, L. et al.: LearnLM: Improving Gemini for Learning (2024). https://doi.org/10.48550/arXiv.2412.16429.
35. Torres, K.M.: Transforming Higher Education With Microlessons. In: Global Perspectives on Micro-Learning and Micro-Credentials in Higher Education. pp. 59–74 (2024).
36. Xu, Z.: AI in education: Enhancing learning experiences and student outcomes. Appl. Comput. Eng. 51, 1, 104–111 (2024).
37. Yan, L. et al.: Promises and challenges of generative artificial intelligence for human learning. Nat. Hum. Behav. 8, 10, 1839–1850 (2024).
38. Yilmaz, R., Yilmaz, F.G.K.: The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. Comput. Educ. Artif. Intell. 4, 100147 (2023).
39. Zdravkova, K. et al.: Unveiling the Impact of Large Language Models on Student Learning: A Comprehensive Case Study. In: 2024 IEEE Global Engineering Education Conference (EDUCON). pp. 1–8 IEEE (2024).
40. Zimmerman, B.J.: Self-efficacy: An essential motive to learn. Contemp. Educ. Psychol. 25, 1, 82–91 (2000).