Tools, Teammates, or Threats? How Pedagogical Reasoning Shapes Novice Instructional Designers' Judgments About AI in Education

Steven Moore[1][0000-0002-5256-0339] and Christine Kwon[2][0009-0001-2825-2280]

[1] George Mason University, Fairfax VA 22030, USA
StevenJamesMoore@gmail.com
[2] Carnegie Mellon University, Pittsburgh PA 15213, USA

Abstract. As generative AI becomes embedded in instructional design, understanding how novice designers reason about AI's pedagogical implications is critical. This study examines how principle-based pedagogical reasoning and AI integration reasoning co-develop in 27 instructional design graduate students across four learning science modules. Students completed paired pre/post constructed responses scored on two analytic rubrics and two categorical codes capturing agency framing and locus of control. Both principle-based pedagogical reasoning ($d = 1.39$) and AI integration reasoning ($d = 1.31$) improved substantially, with AI integration gains more uniform across modules than principle-based gains. Principle-based pedagogical reasoning was significantly associated with AI integration reasoning, with positive cross-dimensional correlations across all rubric criteria. This indicates that the association extends beyond surface familiarity to recognition, strategy, and explanatory components. Students' agency framings shifted significantly toward human-led conceptions, as "Human Leads + AI Supports" nearly doubled (11.5% to 20.6%) while "AI Threatens" decreased (8.7% to 3.7%). However, their perceived locus of control remained comparatively stable. These findings suggest that grounding in learning science principles supports more sophisticated reasoning about AI's pedagogical role and more collaborative conceptions of human-AI partnership in instructional design.

Keywords: AI literacy · Human-AI partnership · Instructional design

1 Introduction

Advances in generative AI have made it possible for people with limited domain expertise to produce polished educational materials in minutes, from lesson plans to quizzes. Education has been quick to adopt these tools, with instructors increasingly using large language models (LLMs) to generate the very assignments that students may also use LLMs to complete [3, 8]. These capabilities have also enabled genuinely novel learning experiences, including instruction that can be personalized in real time and adapted to learners' needs at a scale that was previously difficult to achieve [2]. Yet the ease of producing such educational content does not guarantee the quality of learning it supports. Designing effective instructional materials requires nuanced reasoning about how specific design choices elicit learning processes [14]. This type of reasoning is continually developing for novice instructional designers and new instructors, even as AI amplifies their productivity [13, 4]. This raises a central issue for AI-enabled instructional design. When humans and AI co-produce learning materials, the ease of generating content may outpace designers' ability to apply appropriate learning science principles and reason about AI's pedagogical implications.

Prior research has examined whether generative AI improves the quality of instructional artifacts and whether AI-supported workflows affect learning outcomes [8]. Yet the success of AI-supported instructional design depends not only on the artifacts produced or outcomes observed, but also on the reasoning that guides when and how AI is used to support learning [17]. If these tools promise faster production and higher-quality learning experiences, then the next question becomes what conditions make that promise plausible. One likely condition is that designers possess enough pedagogical content knowledge to evaluate whether AI-generated materials align with established learning science principles, rather than merely appearing well-structured or pedagogically sound on the surface [19]. Without this evaluation, AI may accelerate the production of superficially plausible activities that fail to activate the intended learning mechanisms [9]. Effective human-AI collaboration in instructional design requires designers who can reason about how generative AI intersects with learning science principles and make decisions about how AI-generated suggestions should be adopted, adapted, or rejected.

These considerations motivate the need to characterize how Principle-based Pedagogical Reasoning (PPR) and AI Integration Reasoning (AIIR) develop together in instructional designers. Drawing on existing frameworks for pedagogical content knowledge [22, 28] and AI literacy [18, 19], we distinguish these two competencies central to AI-enabled instructional design. PPR refers to a designer's ability to recognize a learning science principle and justify instructional decisions in terms of that principle's learning mechanisms. AIIR refers to a designer's ability to reason about how generative AI may support or hinder those same mechanisms and to articulate principle-aligned adaptations for using AI in instruction.

Understanding how novices frame AI's role and responsibility in human–AI instructional design work is central to supporting effective collaboration. To address these gaps, we conducted a study that provides evidence on how novice instructional designers' AIIR develops in relation to PPR. We examine students enrolled in a graduate instructional design course that included four learning science modules. Across each module, students completed paired pre- and post-instruction constructed responses. These responses act as observable evidence of two linked competencies, with one focused on applying the given learning science principle in lesson design and a second focused on reasoning about how generative AI might shape learning and design decisions with respect to that same principle. Using descriptive and inferential analyses of these eight paired assessments, we trace how students' reasoning changes after targeted instruction and hands-on experience using generative AI to generate and critique instructional materials. Through this study, we address the following research questions:

RQ1: How do students' principle-based pedagogical reasoning and AI integration reasoning develop from pre- to post-instruction, and do these developmental patterns differ across learning science principles?

RQ2: What is the relationship between novice instructional designers' understanding of learning science principles and their ability to reason about AI's pedagogical impacts, both overall and across specific dimensions of reasoning?

RQ3: How do novice instructional designers conceptualize human-AI agency when reasoning about AI's impact on learning, and do these conceptions shift following instruction?

This work makes three contributions to research on AI-enabled instructional design. First, it provides empirical evidence that novices' PPR is positively associated with the quality of their AIIR. Second, it identifies a developmental pattern in which AIIR improves more uniformly than PPR across learning science principles, suggesting that novices may acquire principle-general ways of evaluating AI's pedagogical role. Third, it characterizes how novice instructional designers conceptualize human-AI agency and control, revealing shifts toward more human-led framings and comparatively stable perceptions of control. These findings inform AIED tool and curriculum design that foreground pedagogical judgment as a foundation for responsible AI integration.

2 Related Work

2.1 Pedagogical Reasoning in Instructional Design

Effective instructional design requires pedagogical content knowledge (PCK), the specialized understanding of how to represent content in ways that make it comprehensible to learners [30, 29]. Recent work has decomposed PCK into declarative (knowledge of facts and principles), procedural (knowledge of instructional procedures), and conditional (knowledge of when and why to apply particular strategies) dimensions [24]. Research consistently shows that novices primarily possess declarative and procedural PCK, while conditional develops only through extensive practice [26]. This developmental trajectory mirrors broader models of expertise acquisition in pedagogy, in which learners progress from acquiring terminology and rules to explaining underlying mechanisms and flexibly responding to novel situations [27, 4]. Even trained teachers have demonstrated PCK and subject matter knowledge that was weakly developed relative to their curriculum design expertise [13], suggesting that PPR remains a persistent challenge even beyond novice stages.

2.2 Generative AI in Instructional Design and Pedagogy

Generative AI tools are increasingly integrated into instructional design workflows for content creation, assessment development, and lesson planning [3, 8].

Beyond the quality of AI-generated content and its effects on learning outcomes, effective integration also demands that designers reason about AI capabilities. Specifically, how they interact with the mechanisms through which learning occurs, in a way that requires expertise to critically evaluate AI output and mitigate risks of over-reliance [9]. Emerging frameworks for AI literacy provide guidance for characterizing such reasoning. Prior work has identified competencies spanning recognition of AI capabilities, understanding of how AI works, and critical evaluation of how AI should be used [18]. UNESCO's AI Competency Framework for Teachers [19] specifies three progression levels of acquire, deepen, and create. Novice designers typically operate at the acquire level, where they are capable of recognizing that AI has implications for instruction, but unable to analyze tradeoffs or propose nuanced design adaptations.

2.3 Human-AI Agency and Locus of Control

As AI becomes embedded in educational and instructional design practice, how humans conceptualize their relationship with AI systems carries significant implications for design decisions and learning outcomes [11, 31]. Theoretical frameworks distinguish between Agency, who is positioned as the primary actor or decision-maker, and Locus of Control, where influence over outcomes is perceived to reside [28, 23]. These constructs can diverge as a designer might frame humans as rightfully leading AI integration (Agency) while also recognizing that AI substantially shapes the resulting decisions (Locus of Control). Research on human-AI collaboration further suggests that users' mental models of AI significantly affect how they interact with and evaluate AI outputs [25]. Common conceptions can range from AI being a passive tool to being directed to an autonomous agent that replaces human functions. Understanding how novice designers conceptualize these relationships is essential for supporting principled AI integration in instructional design [1, 6].

3 Methods

3.1 Study Context and Participants

This study was conducted in a 14-week graduate course on educational technology at a university in the United States (Spring 2024). The course emphasized learning science principles and their application in instructional design. Students entered with prior program coursework and some experience using generative AI for educational tasks. The study focused on four one-week modules: Universal Design for Learning (UDL), Guided Discovery, Fostering Help-Seeking, and Collaborative Learning. Each module followed a common structure where students completed asynchronous e-learning materials (readings/videos/formative assessments) and attended two 80-minute in-person sessions (lecture, discussion, and activities). Outside of class the students also completed bi-weekly projects intended to provide hands-on experience with various pieces of popular educational technology, such as commonly used online course authoring platforms. As part of these activities, students created microlessons incorporating each module's focal learning science principle using a scaffolded four-step process (topic selection, learning objective formulation, assessment creation, and instructional content writing). Following a counterbalanced A/B design, students alternated between using generative AI (ChatGPT) to assist with microlesson creation and creating the microlessons independently. When using generative AI, students were guided to leverage it for drafting and refining lesson components, and to critically evaluate the output against the focal learning science principle's mechanisms. Full details of the microlesson design process, AI usage protocols, and quality evaluation are reported in [21] and [22]. Students also completed a pre-test at the start of each module and a post-test at the end, which form the primary data for this study (Section 3.2).

Participants were 27 master's students in an educational technology and learning sciences program enrolled in a required second-semester course. Students ranged in age from 22 to 36 years and self-reported gender was 21 female and 6 male. To contextualize the label "novice instructional designers", we note that most participants had limited formal teaching or instructional design experience. Students applied learning science principles across self-selected instructional contexts, such as technology, professional development, or cooking, rather than a single content domain. All students reported prior exposure to generative AI tools for general academic purposes and from previous coursework using it for instructional and assessment development.

3.2 Assessment Instruments

Each module included a paired pre-test and post-test consisting of two constructed-response prompts. The pre-test was administered prior to module instruction and the post-test at the end of the module before the course moved to the next topic. Students were prohibited from using generative AI for these assessments. The first prompt elicited PPR by asking students to identify key lesson features and strategies aligned with the focal learning science principle ("What are key aspects of a lesson that incorporates strategies for *[Principle]*?"). The second prompt elicited AIIR by asking students to anticipate how generative AI might shape learning and design decisions with respect to the same principle ("How might the use of generative AI integrate with *[Principle]*?"). Across the four modules and two timepoints, each student contributed eight response occasions (4 modules × 2 timepoints), yielding two open-ended responses per occasion.

Responses were evaluated using two analytic rubrics and two categorical codes, each informed by prior frameworks on pedagogical reasoning and response structure. Rubric A, Principle-based Pedagogical Reasoning (PPR), draws on PCK decompositions [24, 15] and the SOLO taxonomy [5, 7] to capture a three-level progression from everyday recognition of a principle's general idea. This includes partial technical articulation of strategies to integrated understanding of its pedagogical implications. Rubric B, AI Integration Reasoning (AIIR), draws on AI literacy and competency frameworks [18, 19] to capture how students con-

Table 1. Rubric dimensions for evaluating Principle-based Pedagogical Reasoning and AI Integration Reasoning along with verbatim student excerpts presenting low vs. high performance on each dimension.

Dimension	Description	Low Example (1)	High Example (3)
	Rubric A: Principle-based Pedagogical Reasoning		
Principle Recognition	Recognition of core concepts and terminology	"it involves students working together"	"positive interdependence, individual accountability, and team interaction"
Strategy Identification	Identification of instructional strategies	"guide students to resources"	"By providing opportunities for students to self-reflect, you can"
Reasoning Quality	Explanation of purpose, context, and mechanisms	"lets them figure out items by themselves"	"Because instructors are not the focal point of the lesson"
	Rubric B: AI Integration Reasoning		
Principle–AI Connection	Specificity of AI–principle connection	"help evaluate material for UDL"	"adding captions or generating explanations to clarify"
Stance Complexity	Consideration of benefits, risks, and conditions	"could be used in constructing scaffolding to help"	"either aid or hurt the help-seeking, by providing"
Constructive Orientation	Forward-looking implications for practice	"teachers can generate group assignments using"	"creating customized materials based on the abilities of each group"

nect generative AI to the focal learning science principle. Scores progress from generic statements about AI with no principle-specific connection, through partial connection, to nuanced analysis involving multiple principle-linked tradeoffs and constructive suggestions for practice. Because effectively using AI tools is distinct from articulating how AI affects the learning mechanisms of one's designs, Rubrics A and B are scored independently for each response occasion.

Additionally, responses to the AI integration prompt were coded on two categorical dimensions informed by prior conceptualizations of human–AI agency [28, 23]. The first, Agency Framing, characterizes how the response positions the human–AI relationship (e.g. AI as a tool or AI as a threat). The second, Locus of Control, characterizes where decision-making influence is perceived to reside (e.g. AI-directed vs. human-directed). Table 1 summarizes the dimensions of Rubrics A and B, each scored 0-3, with verbatim student excerpts illustrating low and high scores. Table 2 defines the categorical coding schemes for Agency

Table 2. Coding framework for agency framing and locus of control along with verbatim student excerpts illustrating each category.

Category	Description	Example
	Agency Framing	
AI Replaces	AI takes over human functions; humans become passive	"AI could be the instructor and do the lesson"
AI as Tool	AI as instrument humans use; purely instrumental	"Students use it for more accurate feedback"
Human Leads + AI Supports	Human agency primary; AI augments human judgment	"AI can be used to help guide and support the creation"
Partnership	Collaborative framing; shared, complementary roles	"generative AI can consult as a partial expert"
AI Threatens	AI framed primarily as risk to human roles or agency	"introduce biases that instructors might not notice"
	Locus of Control	
AI-Directed	AI makes primary decisions; humans react or accept	"parse through student confused and return to them a solution"
Shared	Distributed decisions; human and AI exercise judgment	"Generative AI can help with identifying where opportunities"
Human-Directed	Humans make decisions; AI provides input or options	"So the user still needs to have oversight"

Framing and Locus of Control with verbatim student excerpts for each category. A complete version of the rubrics and codes used is made available[3].

Across all response occasions, PPR responses averaged 25.8 words ($SD = 22.1$) and AIIR responses averaged 29.5 words ($SD = 21.2$). Post-test responses were significantly longer than pre-test ones for both PPR ($t(26) = 5.82$, $p < .001$) and AIIR ($t(26) = 5.37$, $p < .001$). Response length was strongly correlated with rubric scores at the observation level (PPR $r = .74$, AIIR $r = .61$), which is expected for constructed response measures where more elaborated reasoning necessarily requires more text. Note, we re-estimated key models controlling for word count and results were substantively unchanged.

3.3 Coding Procedures

Two raters with expertise in learning sciences and instructional design independently coded all student responses, blind to student identity and timepoint. Prior to independent coding, raters calibrated by reviewing the rubric and anchor examples, independently coding approximately 10% of responses per module, and discussing disagreements until consensus was reached.

Following calibration, raters independently coded all remaining responses. Inter-rater reliability was computed using weighted Cohen's kappa (quadratic

---

[3] `https://osf.io/z8fhq/overview?view_only=7e55db53e7874ea4add3c63b06074c7b`

Table 3. Inter-rater reliability for the rubric dimensions and categorical tags.

Dimension	**Scale**	κ	**% Exact Agreement**
Rubric A: Pedagogical-based Principle			
A1: Principle Recognition	0-3	.79	81%
A2: Strategy Identification	0-3	.71	72%
A3: Reasoning Quality	0-3	.69	66%
Rubric B: AI Integration			
B1: Principle-AI Connection	0-3	.77	74%
B2: Stance Complexity	0-3	.68	69%
B3: Constructive Orientation	0-3	.70	66%
Categorical Tags			
Agency Framing	1-5	.84	79%
Locus of Control	1-3	.93	86%

weights) for the six ordinal rubric dimensions (A1-A3 and B1-B3, scale 0-3) and unweighted Cohen's kappa for the two categorical codes [10]. The categorical codes required a substantive claim about AI's relationship to the focal principle, responses too vague to reflect a codeable stance ($n = 4$, scoring at the floor on Rubric B) were excluded, resulting in slightly reduced counts for the Agency Framing and Locus of Control analyses. All disagreements were resolved via discussion to produce consensus scores used in subsequent analyses. Disagreements were typically within one score point (adjacent categories), consistent with the use of quadratic-weighted κ for ordinal dimensions. As shown in Table 3, reliability was moderate to substantial across all dimensions.

3.4 Analysis

All analyses account for within-student dependence across modules and timepoints. For each response occasion consisting of each module at each timepoint, rubric criteria scores were summed to produce composite totals forPPR ($A_{\text{Total}} = A1 + A2 + A3$, range 0-9) and AIIR ($B_{\text{Total}} = B1 + B2 + B3$, range 0-9). This yields one A_{Total} and one B_{Total} for every student per module for their pre- and post-test. Student pre-post development was then assessed using two-way repeated-measures ANOVAs with Timepoint (pre, post) and Module (UDL, Guided Discovery, Fostering Help-Seeking, Collaborative Learning) as crossed within-subjects factors. This tests whether scores improved from pre- to post-test and whether those gains varied by module. Effect sizes are reported as generalized eta-squared (η^2_G) for ANOVA effects and Cohen's d_z for within-subject paired comparisons. Additionally, to examine the relationship between the two reasoning competencies, we computed Pearson's r and Spearman's ρ between rubric totals, fit a linear mixed-effects model predicting B_{Total} from A_{Total} with Module as a fixed effect and Student as a random intercept, and computed cross-correlations between rubric subdimensions. Agency Framing and Locus of Control distributions were compared across timepoints using chi-square tests, supplemented by transition matrices and direction-of-change summaries.

4 Results

In this study, we first examine whether students' PPR (Rubric A) and AIIR (Rubric B) improved from pre- to post-instruction across the four learning science modules. We then test whether stronger PPR is associated with more sophisticated principle-linked AIIR for the same learning science principle. Finally, we analyze how students' agency framings of the human–AI relationship shift over time as they gain knowledge of the principles and experience using generative AI in instructional design tasks.

Table 4. Pre-Post test comparison of Principle-based Pedagogical Reasoning and AI Integration Reasoning scores.

Measure	**Pre M (SD)**	**Post M (SD)**	**Change**	**Effect Size**
Pedagogical Reasoning	3.25 (1.74)	5.07 (2.28)	+1.82	$d = 1.39$
AI Integration Reasoning	3.68 (1.72)	5.07 (2.04)	+1.39	$d = 1.31$

4.1 Development of Reasoning

Across the four modules, both PPR and AIIR improved from pre- to post-test with large effect sizes (Table 4). For PPR, there was a significant main effect of Timepoint ($F(1, 26) = 37.06$, $p < .001$, $\eta^2_G = .18$) and a modest main effect of Module ($F(3, 78) = 3.47$, $p = .02$, $\eta^2_G = .04$). This reflects that some principles, such as Guided Discovery, elicited higher overall scores. The Timepoint × Module interaction was not significant ($F(3, 78) = 1.26$, $p = .30$). For AIIR, the main effect of Timepoint was significant ($F(1, 26) = 24.61$, $p < .001$, $\eta^2_G = .13$), but neither the Module main effect ($F(3, 78) = 0.88$, $p = .45$) or the interaction ($F(3, 78) = 0.38$, $p = .77$) reached significance. Module-level paired comparisons confirmed pre-post gains for PPR across all four modules and for AIIR in three of four (Table 5).

These findings indicate that both types of reasoning improved following instruction. PPR in particular showed modest differences by module in overall level, whereas AIIR appeared more uniform across principles in this dataset. This is consistent with the possibility that students were developing a more principle-general way of thinking about AI's pedagogical impacts that develops alongside, but not fully determined by, principle-specific pedagogical knowledge.

4.2 Relationship Between Principle and AI Reasoning

PPR (Rubric A total) and AIIR (Rubric B total) were moderately and positively related (Spearman $\rho = .35$, $p < .001$; Pearson $r = .40$, $p < .001$), corresponding to approximately 12–16% shared variance. In variance terms, this level of shared variance indicates a meaningful relationship for two constructed-response assessments. This relationship remained when accounting for repeated measures within students and controlling for module in a mixed-effects regression ($B\_\text{Total} \sim A\_\text{Total} + \text{Module} + (1|\text{Student\_ID})$): A_Total was a significantly

Table 5. Changes in average pre-post score across Principle-based Pedagogical Reasoning (A) and AI Integration Reasoning (B) collapsed by module.

Module	ΔA_{Total}	d_z	ΔB_{Total}	d_z
Guided Discovery	$+2.58^{***}$	0.90	$+1.62^{**}$	0.60
Fostering Help-Seeking	$+1.85^{***}$	0.81	$+0.85^{\dagger}$	0.37
Universal Design for Learning	$+2.08^{**}$	0.67	$+1.60^{**}$	0.58
Collaborative Learning	$+0.93^{*}$	0.46	$+1.67^{**}$	0.65

Note. Δ indicates Post–Pre change. $^{\dagger}p = .067$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

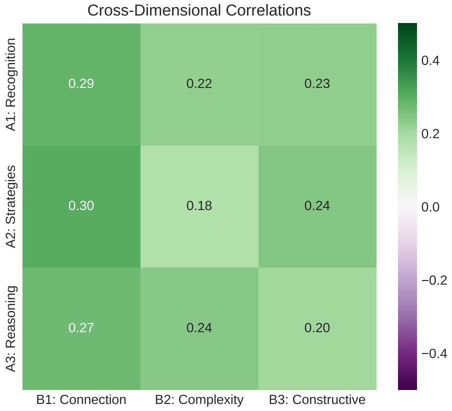

Fig. 1. Spearman correlation heatmap between Rubric A (PPR) and Rubric B (AIIR) dimensions where all correlations significant at p < .01.

associated with B_Total ($\beta = 0.36$, SE = 0.06, $p < .001$). This means for each 1-point increase in PPR corresponded to an average increase of 0.36 points in AIIR (on the 0-9 scale), with no reliable differences in mean B_Total across modules after controlling for A_Total.

To examine whether this relationship was broad or driven by a single subskill, we computed cross-correlations between the three Rubric A dimensions and three Rubric B dimensions (Figure 1). All nine correlations were positive and statistically significant (all $p < .01$), indicating that the association was not limited to one dimension. Three theoretically meaningful pairings were particularly notable: (1) **Principle Recognition (A1)** correlated with **Principle-AI Connection (B1, $\rho = .29$)**, suggesting that clearer identification of the learning principle supports more Principled-based Pedagogical Reasoning about AI, rather than generic AI statements); (2) **Strategy Identification (A2)** correlated with **Constructive Orientation (B3, $\rho = .24$)**, suggesting that knowledge of instructional strategies supports more actionable recommendations for adapting practice in the presence of AI; and (3) **Reasoning Quality (A3)** correlated with **Stance Complexity (B2, $\rho = .24$)**, suggesting that more mechanistic or conditional pedagogical reasoning aligns with more nuanced consideration of AI tradeoffs.

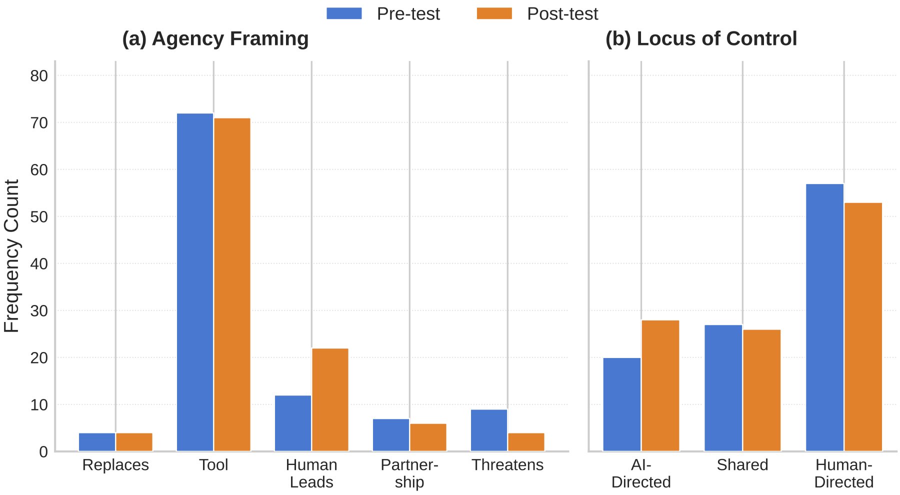

Fig. 2. Distribution of Agency Framing (a) and Locus of Control (b) codes in pre- and post-instruction responses, collapsed across modules.

Together, these findings indicate that students who demonstrate stronger PPR also produce more sophisticated reasoning about generative AI's pedagogical impacts. Importantly, the relationship operates across recognition, strategy, and explanatory components, suggesting that deep pedagogical knowledge, not just surface familiarity with terminology, supports thoughtful AIIR.

4.3 Agency and Control Shifts

For *Agency Framing*, *AI as Tool* was the dominant conception at both timepoints (69.2% pre, 66.4% post) with the counts shown in Figure 2a. Despite this stable modal category, the overall distribution shifted significantly from pre to post, $\chi^2(16) = 32.20$, $p = .009$. The most notable changes were an increase in *Human Leads + AI Supports* (11.5% → 20.6%) and a decrease in *AI Threatens* (8.7% → 3.7%), a majority shift towards more human-led framings. Together, these patterns indicate that post responses were more likely to position AI in a supportive role under human direction and less like a threat.

Locus of Control showed a different pattern (Figure 2b). The overall distribution did not shift significantly, $\chi^2(4) = 9.43$, $p = .053$, with *Human-Directed* remaining the most common code at both timepoints (54.8% pre, 49.5% post). Descriptively, however, there was a modest increase in *AI-Directed* control (19.2% → 26.2%) and module-level patterns suggested heterogeneity in these shifts.

5 Discussion

This study examined how novice instructional designers' Principle-based Pedagogical Reasoning (PPR), AI Integration Reasoning (AIIR), and conceptions of human-AI agency develop through instruction on learning science principles. First, students improved substantially from pre to post on both rubrics, with large within-subject effects for PPR and AIIR. These gains were broadly consistent across all four modules. Second, PPR was positively associated with AIIR and this relationship held across recognition, strategy, and explanatory dimensions. Third, students' agency framings shifted towards a more human direction, as "Human Leads + AI Supports" nearly doubled (11% to 21%) while "AI Threatens" decreased. These findings provide insights into how students developed more differentiated judgments about AI's pedagogical role and the human-AI relationship in educational contexts.

The significant positive association between students' PPR and their reasoning about generative AI's pedagogical implications has direct implications for how we prepare instructional designers to work with AI. A common refrain among students and practitioners is "Why do I need to learn this if I can just ask an LLM?" [12]. Our findings offer an empirical response, as we demonstrate that students who exhibited stronger PPR also produced more principle-connected judgments about generative AI. This pattern aligns with the PCK [30] and related frameworks (e.g. TPACK [14]). These emphasize that effective design requires more than knowing content or tools, but also understanding how instructional choices connect to learning processes in context. We extend this logic to human-AI collaboration, where designers need principled pedagogical knowledge not merely to use AI tools, but to evaluate whether AI-generated suggestions align with how learning actually works. Without this foundation novices may accept AI output that appear plausible yet conflicts with established learning principles, limiting the quality and safety of AI-supported instructional decisions.

In the present study, novices' conceptions of human-AI collaboration become more nuanced with instruction and experience, but not uniformly across constructs. Students' role framings (Agency) shifted significantly toward more human-led conceptions, as "Human Leads + AI Supports" increased, while "AI Threatens" decreased. This suggests that role conceptions are malleable and responsive to learning experiences. At the same time, perceived control over decisions (Locus of Control) was comparatively stable and trended toward greater acknowledgment of AI-directed influence. Students may simultaneously endorse humans being in command while recognizing that AI can shape decisions in practice through persuasive outputs or efficiency affordances [20]. This divergence is part of a theoretically important separation between who should lead and what actually steers choices. These results suggest that the combination of principled instruction and structured AI experience is associated with novices moving away from simplistic replacement/threat framings and toward a more collaborative mental model of AI as a teammate.

These results suggest novices may develop principle-general ways of evaluating AI even when principle-specific pedagogical understanding remains uneven. Instructional design curricula may therefore benefit from foregrounding learning mechanisms and requiring students to justify AI use (or non-use) in terms of those mechanisms, not just productivity. Structured reflection checkpoints that force explicit human decisions may reduce uncritical adoption and support principled "AI-in-the-loop" design habits. As generative AI becomes routine in instructional workflows, responsible integration requires a foundation in learning science principles. The role of the instructional designer will not become obsolete as generative AI continues to grow, but rather be enhanced by such advancements [16]. Instructional designers will therefore need to evaluate AI outputs against learning mechanisms and to adapt them in ways that facilitate learning rather than hindering it.

This study is not without limitations, as it involved 27 students from a single graduate program at one U.S. institution, constraining generalizability. Our constructed response measures capture articulated reasoning, and although two trained raters achieved acceptable reliability, rubric evaluation necessarily involves some subjectivity. Because instruction and hands-on AI experience were combined within an authentic course, the design does not isolate their individual contributions to the observed gains. This limits causal specificity but offers an ecologically valid picture of how PPR and AIIR co-develop within realistic instructional design training. Although the non-significant Timepoint × Module interaction suggests that pre-post gains did not differ systematically by module, we cannot rule out the possibility that cumulative experience with the course structure or generative AI tools contributed to gains observed in later modules. Despite these constraints, the consistency of findings across all four modules strengthens confidence in the observed patterns.

6 Conclusion

This study set out to examine whether novice instructional designers can develop sophisticated reasoning about AI's integration with instructional design, and whether such reasoning depends on foundational pedagogical knowledge. Our findings affirm these aims as we found learning science principle understanding was significantly associated with AI Integration Reasoning, and both capacities improved substantially across modules that combined targeted instruction with hands-on design experience. Students' conceptions of human-AI agency shifted toward more human-centered framings, suggesting that appropriate mental models of AI collaboration can be cultivated alongside pedagogical expertise. Beyond documenting these patterns, this work provides an account of what responsible AI integration demands cognitively. Our results are consistent with the view that this transition requires more than exposure to AI capabilities, suggesting that grounding in learning science principles may be what allows humans to evaluate and appropriately constrain AI's role in instruction. As generative AI becomes increasingly embedded in educational practice, ensuring that designers possess both deep pedagogical knowledge and the capacity to reason critically about AI's impacts may be essential for realizing human-AI partnerships that genuinely serve learning. One in which tools scaffold designers to articulate mechanisms, anticipate tradeoffs, and enable generative AI to function as a true collaborator that strengthens instructional design.

Acknowledgments. We thank the Eberly Center 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 tothe content of this article.

References

1. Aleven, V., Mavrikis, M., McLaren, B.M., Nguyen, H.A., Olsen, J.K., Rummel, N.: Six instructional approaches supported in aied systems. Handbook of artificial intelligence in education pp. 184–228 (2023)
2. Arslan, B., Wilschut, T., van Rijn, H., van der Velde, M.: Evidence-based dynamic personalization for learning and assessment tools (ed-plat): Machine-and learner-driven adaptation to support all learners. In: International Conference on Artificial Intelligence in Education. pp. 477–491. Springer (2025)
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. Education Sciences **13**(10), 998 (2023)
4. Berliner, D.C.: Describing the behavior and documenting the accomplishments of expert teachers. Bulletin of science, technology & society **24**(3), 200–212 (2004)
5. Brabrand, C., Dahl, B.: Using the solo taxonomy to analyze competence progression of university science curricula. Higher Education **58**(4), 531–549 (2009)
6. Caskurlu, S., Yalçın, Y., Hur, J., Shi, H., Klein, J.D.: Data-driven decision-making in instructional design: Instructional designers' practices and strategies. TechTrends pp. 1–12 (2025)
7. Chan, C.C., Tsui, M., Chan, M.Y., Hong, J.H.: Applying the structure of the observed learning outcomes (solo) taxonomy on student's learning outcomes: An empirical study. Assessment & Evaluation in Higher Education pp. 511–527 (2002)
8. Choi, G.W., Kim, S.H., Lee, D., Moon, J.: Utilizing generative ai for instructional design: Exploring strengths, weaknesses, opportunities, and threats. TechTrends **68**(4), 832–844 (2024)
9. Cipriano, B.P., Alves, P., Denny, P.: A picture is worth a thousand words: Exploring diagram and video-based oop exercises to counter llm over-reliance. In: European Conference on Technology Enhanced Learning. pp. 75–89. Springer (2024)
10. Fleiss, J.L., Cohen, J.: The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educational and psychological measurement **33**(3), 613–619 (1973)
11. Fourie, K., Goto, J., Lautenbach, G.: Unlocking the potential of large language models in education: Factors influencing adoption by instructional designers and academics. Journal of Information Technology Education: Research **25**, 02 (2026)
12. Grande, V., Kiesler, N., Francisco R, M.A.: Student perspectives on using a large language model (llm) for an assignment on professional ethics. In: Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1, pp. 478–484. ACM New York, NY (2024)
13. Huizinga, T., Handelzalts, A., Nieveen, N., Voogt, J.: Fostering teachers' design expertise in teacher design teams: conducive design and support activities. Curriculum journal **26**(1), 137–163 (2015)
14. Koehler, M., Mishra, P.: What is technological pedagogical content knowledge (tpack)? Contemporary Issues in Technology and Teacher Education **9**(1), 60–70 (2009)

15. Kramarski, B., Michalsky, T.: Preparing preservice teachers for self-regulated learning in the context of technological pedagogical content knowledge. Learning and instruction **20**(5), 434–447 (2010)
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 Learning **28**(3), 207–231 (2024)
17. Li, H., Fang, Y., Zhang, S., Lee, S.M., Wang, Y., Trexler, M., Botelho, A.F.: Arched: A human-centered framework for transparent, responsible, and collaborative ai assisted instructional design. In: Innovation and Responsibility in AI-Supported Education. pp. 94–104. PMLR (2025)
18. Long, D., Magerko, B.: What is ai literacy? competencies and design considerations. In: Proceedings of the 2020 CHI conference on human factors in computing systems. pp. 1–16 (2020)
19. Miao, F., Shiohira, K., et al.: AI competency framework for students. UNESCO Publishing (2024)
20. Mills, S., Sætra, H.S.: The autonomous choice architect. AI & society **39**(2), 583–595 (2024)
21. Moore, S., Eckstein, L., Kwon, C., Stamper, J.: Generative ai in instructional design education: Effects on novice microlesson quality. In: International Conference on Artificial Intelligence in Education. pp. 464–471. Springer (2025)
22. Moore, S., Eckstein, L., Kwon, C., Stamper, J.: Integrating generative ai into instructional design practice: Effects on graduate student learning and self-efficacy. In: European Conference on Technology Enhanced Learning. pp. 367–381. Springer (2025)
23. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans **30**(3), 286–297 (2000)
24. Park, S., Oliver, J.S.: Revisiting the conceptualisation of pedagogical content knowledge (pck): Pck as a conceptual tool to understand teachers as professionals. Research in science Education **38**(3), 261–284 (2008)
25. Rezwana, J., Maher, M.L.: An exploration of mental models of ai in human–ai co-creativity: A framework and insights. ACM Transactions on Interactive Intelligent Systems **15**(4), 1–26 (2025)
26. Saks, K., Ilves, H., Noppel, A.: The impact of procedural knowledge on the formation of declarative knowledge: How accomplishing activities designed for developing learning skills impacts teachers' knowledge of learning skills. Education Sciences **11**(10), 598 (2021)
27. Selinger, E.M., Crease, R.P.: Dreyfus on expertise: the limits of phenomenological analysis. Continental Philosophy Review **35**(3), 245–279 (2002)
28. Shneiderman, B.: Human-centered ai: A new synthesis. In: IFIP Conference on Human-Computer Interaction. pp. 3–8. Springer (2021)
29. Shulman, L.: Knowledge and teaching: Foundations of the new reform. Harvard educational review **57**(1), 1–23 (1987)
30. Shulman, L.S.: Those who understand: Knowledge growth in teaching. Educational Researcher **15**(2), 4–14 (1986)
31. Zimmerman, A., Janhonen, J., Beer, E.: Human/ai relationships: challenges, downsides, and impacts on human/human relationships. AI and Ethics **4**(4), 1555–1567 (2024)