RHTTE

RESEARCH HIGHLIGHTS IN
TECHNOLOGY AND TEACHER
EDUCATION

Connecting Theory and Practice: Large Language Models as Tools for PCK Development in Teacher Education

Suparna Chatterjee
New Mexico State University
suparna@nmsu.edu

Melissa Warr
New Mexico State University
warr@nmsu.edu

Abstract:
Large language models (LLMs) are adept at adopting personas and afford unique approaches for supporting teaching and learning. In this article, we present the first stage of an Educational Design Research study investigating the questions 1) What design principles support a process for the development of PCK in teacher candidates using LLMs? 2) What types of thinking and connections occur through this practice? The purpose of this research is to (1) design a use of LLMs that supports teachers in the development of pedagogical content knowledge (PCK), and (2) develop a theoretical understanding that describes how learning occurs through the design. Initial findings emphasize three themes: the uniqueness of the LLM, the role of reflection in learning with LLMs, and the importance of scaffolding. These findings support three design principles, emphasizing how the use of structure, modeling, and reflection can support effective role-play interactions with LLMs.

Introduction

There has been much excitement about the new accessibility to generative AI tools, particularly large language models (LLMs). ChatGPT has been described as an "around-the-clock educator," changing perceptions about teaching and learning within six months of its launch at the end of 2022 (Wu, 2023, para. 1). While much of the conversation has focused on uses for writing, personalized learning, and reducing teacher workload (Herft, 2023; Chan & Hu, 2023), technologies come with specific affordances that sometimes offer possibilities for shifting pedagogies that support new types of learning (Wu, 2024). In this article, we explore whether LLMs offer a new and effective way to help future teachers develop pedagogical content knowledge (PCK) through role-play activities. We describe our initial experiment with this approach, explain how it supported the development of reflective practice in a unique way, and provide suggestions for implementation and research.

PCK—knowledge at the intersection of pedagogy and content (Shulman, 1986)—is critical to effective teaching (Sarjoni et al., 2020). Developing PCK traditionally requires classroom experience and structured reflection (Agricola et al., 2020). However, given the recent widespread availability of LLMs such as ChatGPT, we wondered whether these models could support the development of teacher candidates' PCK. In this research, we asked ChatGPT to be a third-grade student with misconceptions in math and science (explained in detail below). Misconceptions can be described as thoughts, notions, and perspectives from life experiences and informal learning that lead to incorrect interpretations of phenomena (Soeharto et al., 2019). We practiced teaching the "student," reflected on our own experiences, and reviewed each other's interactions. We were particularly interested in the dynamics of these role-play interactions, including our own thinking and learning process and how our PCK developed through the activity. We developed design principles—guidelines for design that connect to theoretical understanding (Bakker, 2018)—that would help us design applications for teacher candidates. Thus, the purpose of this research is twofold: first, to design an application of LLMs with associated design principles that aids PCK development, and second, to build theoretical knowledge about the mechanisms of this type of learning with LLMs. In the next section, we provide background to four core ideas in our work: LLMs, PCK, reflection, and scaffolding.

Literature Review

Large Language Models in Education

We learn language by interacting with other language users in a community; thus, language is embodied through interactions (Shanahan et al., 2023). On the other hand, LLMs are disembodied neural networks trained on a huge amount of human-generated text and trained to

predict the next word or word part (token) given a sequence of tokens (Shanahan et al., 2023). The ability to generate natural language dialogue can be utilized for interactive pedagogical engagement (Bahrini et al., 2023).

Jeon and Lee (2023) discussed four ChatGPT roles for educators—interlocutor, content provider, teaching assistant, and evaluator—highlighting the importance of teachers' pedagogical skills when utilizing AI tools. Recent use of generative AI in education includes content creation (e.g., Trust et al., 2023), personalized learning and differentiation (e.g., Baidoo-Anu & Anash, 2023), and evaluation (e.g., Baido-Anu & Anash, 2023). For example, a recent survey by Microsoft (2024) reported that the most common use of AI by teachers is for creating lessons and assignments. Khan Academy's Khanmigo attempts to provide personalized tutoring to students (Singer, 2023). Finally, Texas recently announced it will use AI as an initial grader in standardized tests (Peters, 2024).

Less common is the use of LLMs for instructional practice, such as through role-play interactions. Many surveys of how educators use LLMs do not even include this option (e.g., Microsoft, 2024; Open Innovation Team and Department for Education, 2024). However, given LLMs facility with creating personas (Schuller et al., 2024), this may be an effective practice. Next, we describe one area where role-plays in LLMs may be particularly powerful: for developing PCK.

Developing Pedagogical Content Knowledge

Shulman (1986) introduced pedagogical content knowledge (PCK) and distinguished it as one of the seven categories of teachers' knowledge, including content, general pedagogical, curricular and knowledge of learners, educational contexts, and educational goals. Added to these six areas is pedagogical content knowledge, which Shulman (1987) defined as "that special amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding" (p. 8). PCK informed the later development of the Technological, Pedagogical, and Content Knowledge (TPACK) framework commonly used in educational technology research (Mishra & Koehler, 2006). However, although this study is related to TPACK in that we are exploring how technology can help teacher candidates develop PCK, the focus of knowledge development for teacher candidates is specifically PCK. It includes subject matter knowledge as well as how to help others learn that knowledge, such as appropriate sequencing of subject matter and identifying and addressing common misconceptions. The role-play approach we present may support teacher educators' TPACK, but that is beyond the scope of this study.

A PCK-informed approach supports teachers in creating learning activities and translating their subject-matter expertise into "forms and representations that are appropriate for students' developmental stages and relevant to them (Sarkar, 2024, p. 1). Carlson et al. (2019) described the PCK Refined Consensus Model (RCM), "complex layers of knowledge and experiences" that direct educators' professional activity to influence students' learning (p. 82) and is divided into three realms: collective PCK, personal PCK, and enacted PCK. Collective PCK (cPCK) is widely accepted knowledge regarding teaching and learning particular subjects that has been developed from research and/or collective wisdom; personal PCK (pPCK) describes a

teacher's individualized teaching knowledge and skills about teaching and learning a particular topic; and enacted PCK (ePCK) is what a teacher applies in a particular teaching situation, including planning for and reflection on teaching (Alonzo et al., 2019; Carlson et al., 2019).

Experience in classroom settings, such as during student teaching, clinicals, or internships, provides pre-service teachers with opportunities to put their knowledge into practice (Mazibe, 2024). Nilsson and Karlsson (2019) found recorded instructional episodes to be an effective instrument for PCK development, allowing for later reflection and fostering the growth of PCK. Nilsson and Karlsson (2019) emphasized reflection-in-action, the capacity to create and use knowledge during teaching, and reflection-on-action, making meaning of past experiences to improve future action (Schön, 1983), as crucial in developing PCK.

Teachers refine skills through practice (Mikeska et al., 2023). Authentic practice such as teacher candidates' participation in live role play and rehearsals with peers (Benedict-Chambers et al., 2020) or live and digitally simulated teaching (Mikeska & Howell, 2020) create opportunities to engage in practices and develop PCK. In this article, we describe research that builds on past work of PCK development through the application of LLMs. This approach offers an innovative tool to practice and develop PCK, not only overcoming the challenge of the limited time and experience teacher candidates spend in the actual classroom but also supporting unique affordances that more effectively scaffold reflection.

Reflection and Scaffolding

A core component of developing practitioner knowledge is reflection (Schön, 1983; Arefian, 2023). Dewey discussed reflective action as being open minded, taking responsibility (being open to the consequences of actions), and engaging in wholeheartedness. (Zeichner & Liston, 1993). Dewey explained, "[Reflection] emancipates us from merely impulsive and routine activity … enables us to direct our actions with foresight and to plan according to ends in view of purposes of which we are aware. It enables us to know what we are about when we act" (Zeichner & Liston, 1993, p. 11).

Perhaps more well-known is Schön's (1983) work, which defined two types of reflection common throughout practitioner work: reflection-in and reflection-on-action. Reflection-in-action describes reflection bound by the "action-present", when "action can still make a difference to the situation" (Schön, 1983, p. 62). Reflection-in-action reflects Dewey's (1993) concept of being open to the consequences of actions—it is through constantly adjusting our actions in response to our actions' results that allows us to be effective in a specific situation. Reflection-on-action describes looking back on an event and considering what was and could have been. Although this reflection cannot directly change the past situation, it can inform future interactions.

The power of reflection is its ability to develop a personal and context-embedded type of knowing, the type needed in complex practices such as teaching (Schön, 1983). Ultimately, "The unique and uncertain situation comes to be understood through the attempt to change it, and changed through the attempt to understand it" (Schön, 1992, p. 132). It is this active knowledge building that can help teachers be effective in a constantly changing context.

Although reflection-in and on-action are critical to developing practitioner knowledge, teacher candidates often need support to do so effectively (Umutlu & Kim, 2020). Scaffolding can provide this support. Bruner (1970) conceptualized scaffolding reflecting the fundamental idea of Vygotsky's (1978) Zone of Proximal Development, which asserts that the individual learner creates meaning through their interpersonal interaction. The teacher, as a facilitator, encourages students' active participation and inquiry, promoting in-depth comprehension. Initially, learners depend on the support of the teacher as they develop new knowledge and skills. Gradually, learners become independent as their thought processes and skills develop, resulting in a commensurate reduction in the assistance they need to be successful. Pedagogical scaffolding is characterized by providing provisional support and subsequent removal of it, emphasizing the gaining of relevant knowledge and skills by the learners.

The function of LLMs—specifically, ChatGPT—in educational contexts may be considered a scaffold, supporting students to acquire new skills and improve comprehension (Wu, 2024). Or, as we describe in this article, LLMs may help teacher educators scaffold student learning and reflection. Next, we provide an overview of the methodology we are using to explore this possibility.

Methodology

The purpose of this research is twofold: (1) design a use of LLMs that supports teachers in the development of PCK, and (2) develop theory that describes how learning occurs with LLMs. The theoretical framework in development aims to move beyond simply describing how learning with LLM occurs; it also strives to be "instrumental" (Warr et al., 2020), suggesting guidance for how to design instructional uses of generative AI tools. Thus, the central research questions for this line of work are:

1. What design principles support a process for the development of PCK in teacher candidates using LLMs?
2. What types of thinking and connections occur through this practice?

Given our dual focus on both an intervention and theoretical understanding, we applied educational design research (EDR), a methodology described by McKenney and Reeves (2020). EDR is a type of design-based research (DBR), a methodology developed by Brown (1992), Barab and Squire (2004), Cobb et al. (2003), Sandoval (2014), and Bakker (2018). Researchers who apply EDR use an iterative process to develop both an intervention and theoretical knowledge. Figure 1 provides McKenney and Reeve's (2020) generic model of EDR, illustrating the iterative development of both the intervention and theory.

Figure 1
Generic Model for Conducting Educational Design Research
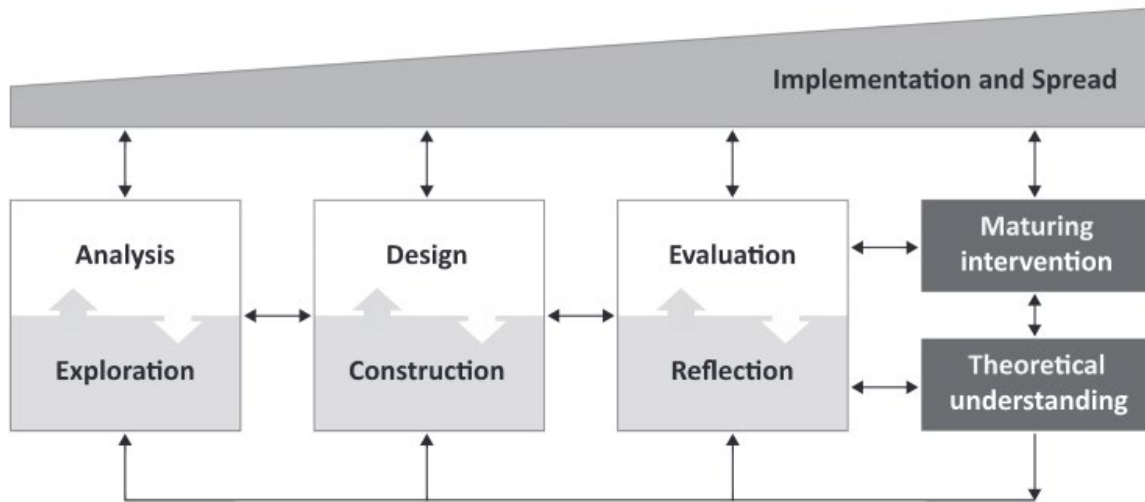
Note. From McKenney & Reeves (2020). Used with permission of the authors.

EDR studies commonly require several stages that build on each other for the development of an intervention and theory. For example, in the research described here, Stage 1 focused on analysis and exploration: researcher exploration and analysis of role play with LLMs to develop PCK. Future research includes Stage 2, where we take what we learned in Stage 1 to design a workshop for teacher educators that explores both the intervention and the theoretical structures that support it. Stage 3 includes evaluating the teacher educator workshop through the data collected, reflecting on the hypothesized design principles, and further refining both the intervention and the theoretical understanding. We then will iterate on our process, returning to design and construction to create activities specifically for teacher candidates, implementing these activities, and engaging in further evaluation and reflection. Through these iterative steps, both the intervention and the theory are developed in the social context of an actual implementation, increasing consequential validity, or the usefulness of the theory (Barab & Squire, 2004).

In this article, we focus on the first stage of the broader EDR study: the initial exploration and analysis of using role play with LLMs for the development of PCK.
Guiding questions for this stage are:

1. Intervention Exploration: What are effective prompts that support the role play activity? What is the experience of engaging in this activity? What are the strengths and weaknesses of this approach? What design principles might support success?
2. Theoretical Understanding: What types of cognitive processes does the participant engage in during this activity?

To develop a deep understanding of the experience and process of this activity, we conducted an initial exploration using self-study; we experimented with the design and used collaborative reflection to refine the process and initiate theory development. This methodology has been found useful in DBR studies to provide reflection on learning throughout the design process (Parsons & Hjalmarson, 2017). The designer is not only creating the intervention but also reflecting on how learning occurs during the intervention. Parsons and Hjalmarson highlighted the power of critical collaboration, where researchers provide one another with feedback and

analysis, furthering the depth of the analysis. We applied this approach to refine our design and deepen our understanding of its dynamics.

Participants

In Phase 1 of the EDR study, the two authors served as researchers and participants. We are faculty at a land grant university in the Southwest Borderlands of the United States. We each have a background in design and educational technology. Suparna (1st author) has experience teaching science, and Melissa (2nd author) mathematics.

Limitations

As our first exploration of using LLMs in role plays to develop PCK in teacher candidates, this study was meant only to provide an initial design and theoretical model to be further refined through research with teacher educators and teacher candidates. Although the narrowness of this study may be seen as a limitation, it is the iterative development of the design and internal reflection on experiences that provide the foundation for future work. Future research will expand this work to a broader range of participants.

Data Collection

Data for Phase 1 of the EDR study included both design artifacts and researcher reflections, akin to self-study DBR as described by Parsons and Hjalmarson (2017). The data consisted of screen recordings of researcher interactions with LLMs, chat transcripts from the interactions (including prompts), and written researcher reflections. As each of us engaged in the role play experiments, we documented our process through recording the screen we were interacting on and stating our thoughts aloud (Koro-Ljunberg et al., 2012; Orderman et al., 2021). This technique allowed us to capture the moment-by-moment reactions we had throughout the activity, including the questions we asked ourselves and our decision-making process on how to interact with the LLM. After completing our interactions, we wrote brief reflections about our experiences.

Data Analysis

We analyzed the screen recordings and reflections independently and collaboratively. First, we watched each other's recordings and reflections, writing comments and reflective memos in response to our research questions (Saldana, 2016). Then we reviewed our own recordings and reflections, continuing to write memos on what we were observing. Next, we discussed our observations, including both the challenges and benefits of the activity design and our own experiences and thought processes. This combined analysis—of both the design and the internal thinking process—led to the findings we describe next.

Findings

In this section, we describe both the design and theoretical principles we developed through our analysis. The initial design consisted of two prompts (see Appendix), pasting these prompts into ChatGPT 3.5 and 4, and interacting with the model. We attempted to teach the LLM "student," focusing on guiding its "understandings" of fractions and astronomy and, at the

conclusion of the interaction, requesting feedback. The appendix contains more detail on each of our interactions.

　　After analyzing our data, we refined the design, which we will describe in the discussion section. Overall, our analysis of this design highlighted three primary themes: that the LLM was different from a "real" student but presented some characteristics that may be seen as affordances of the tool, how we utilized reflection to deepen our knowledge, and the need for scaffolding to effectively support teacher candidates in these interactions. In other words, we found that the affordances of the LLMs support reflection, but the reflection requires scaffolding. Table 1 connects our experience to the theoretical themes, resulting design principles, and implications for future iterations. These are elaborated in more detail below.

Table 1

Connection between theme, design principle, and context for next iteration

Analysis	Theme (Theoretical)	Resulting Design Principle	Implications for Next Iteration
The LLM responses include age-appropriate explanations, with a clearer description of its "thinking process" than that of a typical student. It also introduces new terms and examples. The feedback could be generic or more specific	**Uniqueness of LLM** Enhanced thought process, improved content explanation (language and examples), ability to generate feedback	Encourage critical thinking and mindful reflection on responses, considering the difference between the LLM and a "real student" and noticing the types of language and examples it uses	**Structure** Discuss between interactions; do in groups; ask for specific feedback
During our interactions, the ability to pause between interactions led to careful consideration of questions and responses and also allowed us to return to the transcript for further investigation. Engaging in these role plays involves very limited risk and cost; and enables thinking between questions. We also benefited from reflecting back on our interactions with the LLM	**Reflection in- and reflection on-Action** Our interactions revealed our own reflection in- and on-action	Emphasize reflection during and after interactions with LLM	**Reflection** Structure group activity to have discussions between each interaction and to return to the transcript for reflection after the interaction
Our thinking processes revealed how we carefully considered questions to ask and drew upon unique words or terms the students used; this is unlikely to be the case for	**Scaffolding** Teacher candidates may need scaffolding to support productive interactions	Support careful construction of questions and deconstruct responses	**Modeling** Teacher educators should model their own thinking, gradually giving the teacher candidates

teacher candidates	more responsibility in forming questions and evaluating responses

Uniqueness of LLM

Although many teacher educators use peer-to-peer role play to practice teaching and for PCK development, there are several unique characteristics of doing so with an LLM. The uniqueness is not necessarily better or worse than peer-to-peer role play, but the technology offers unique affordances. We focused on three unique elements of practicing with an LLM: enhanced thought, improved content explanation, and the ability to generate generic and specific feedback.

First, throughout our conversations, we noticed that the LLM seemed much more effective at expressing its thoughts and ideas than a typical 3rd grade student would. For example, the LLM told Melissa:

> When you multiply, like with whole numbers, it's like saying you have a bunch of something. Like if you have 2 times 3, it's like you have 2 groups of 3 so 6 things. But with fractions, it's confusing because how can you have half of something three times? It's not whole things you're talking about...So in my mind, mixing fractions by adding makes more sense

The ability of the LLM to explain its confusion of mixing groups, halves, and wholes, even describing what happens "in my mind," may be beyond what a typical 3rd grader would articulate. However, this doesn't mean it may not be an assumption a real 3rd grader would hold. Thus, this advanced articulation could be both a weakness and a strength. It may give a false impression of how students would react. However, it also might provide insight into the unarticulated thoughts of a typical student.

This leads to our second type of uniqueness—improved content explanation. The LLM often provided a clearer explanation of the content itself, again better than a typical learner, but did so using child-friendly language and terms. It also provided some concrete examples that helped it understand (or misunderstand) the concept. This process could provide future teachers with new ideas for providing simple explanations and examples that they can draw on in their future teaching experiences. It also pushes the instructor to think of the content from different angles and even enhances content knowledge.

Finally, one of the greatest benefits of this practice over regular peer-to-peer conversations is the potential for receiving feedback at the end of practice. In our examples, we noticed that initial feedback seemed quite generic, often getting the same feedback every time such as to use more visuals or examples. However, probing for more detail through specific requests for improvement led to higher quality. Melissa found that when prompted, the LLM was able to suggest specific questions that might have been effective during the conversation as well as identify words and phrases that may have been confusing for multilingual learners.

Reflection

One of the affordances of this exercise was the ability to pause and think between each interaction. We noticed how we thought carefully throughout, including clarifying our own content knowledge to appropriately respond to the student. The unlimited time provided an ideal context for reflection-in-action: we could choose our next move and then think about the results we got from that move. In fact, we could even go back to a previous part of the interaction and change our response, considering how the conversation may play out differently depending on our responses.

We also observed how returning and reflecting on the conversation transcript led to new observations and deeper understandings. This reflection-on-action deepened our PCK, highlighted bias in the AI, and informed assumptions we may have been making in our interactions. For example, when Melissa returned to review her recording, she noticed the types of questions she was asking, and which were more or less effective. She considered how a less effective question could have been improved. Suparna noticed that it provided opportunities to better understand student's thought process by asking "Can you tell me more about…" based on which it elaborated. Melissa also realized something she had not noticed during the activity— that the AI student tended to use cake or pizza when talking about fractions, examples that were not necessarily culturally responsive. Observing this behavior helped her think more critically about what it means to be culturally responsive and the potential biases in AI. Finally, Melissa noticed that she was accepting the AI feedback automatically as if it were "true," even though some of it was quite generic and not necessarily relevant. This observation helped her refine how she interacted with AI tools.

Not only does reflection help teachers develop and mold their PCK and reflection skills, but it is also critical to using AI in general. As previously discussed, the AI provided clearer explanations than a typical student would, and it is important that this is highlighted for teacher candidates. They can observe the types of words the AI uses and the examples it gives and may choose to use the age-friendly vocabulary in future interactions, but they must also be prepared for students who are less articulate. Scaffolding, which we turn to next, can make this interaction effective.

.

Scaffolding

The uniqueness of this exercise, including the ability to stop at any time and discuss questions and responses, enhanced opportunities for reflection-in-action. Furthermore, the transcript of the entire role-play provided an artifact that could support reflection-on-action. However, it is unlikely that teacher candidates will automatically engage in reflection, particularly the type that supports the development of PCK, without scaffolding and modeling.

Role play using the LLM offered the benefit of having time to think through each response, attempting to understand the student's comments, and coming up with an appropriate question to continue the conversation, which improved our learning experience. As we came to the practice with more experience than a typical teacher candidate, we wondered how our

experience impacted our interactions. For example, we picked up on specific terms in our specialty content areas, something that required previous expertise. Melissa noticed the use of the word "whole," and Suparna noticed "observed." However, the significance of these words was not as clear outside of our specialty areas. For example, Melissa questioned whether she would have realized the significance of "observed" given her limited science PCK.

We hypothesized that scaffolding may be critical in supporting this type of reflection with future teachers, perhaps practicing these role plays with a class and small groups before engaging individually, emphasizing careful reflection throughout the process. Guiding reflective thinking about the choices made during the simulation may help novice teachers critically analyze their teaching methods. It could provide insights into factors aiding teacher interaction and engagement in a continuous cycle of activity, feedback, and improvement, which could support contextualizing preservice teachers' PCK development.

Discussion

Applying PCK-focused role-play techniques helped us identify the benefits and limitations of this approach and establish guidelines, or design principles, for how best to engage teacher candidates. The uniqueness of the interaction supported reflection, but effective reflection needs to be scaffolded by teacher educators. We emphasize three practices for future iterations: reflect, structure, and model.

Teachers, by reflecting on their teaching experiences, can develop more nuanced and effective approaches to teaching a subject matter. One of the main affordances of role-play interactions with LLMs is how they can support reflection. Schön's (1986) reflection-in-action calls for reflecting while a difference can still be made in a situation and then reflecting on the result. This can be difficult in a classroom, where interactions are fluid and on-going. However, when working with an LLM, as much time as needed can be taken between each interaction, providing opportunities for deep reflection-in-action. Furthermore, this reflection-in-action can be scaffolded by teachers, who can guide developing teachers' thinking to refine their PCK.

In addition to supporting reflection-in-action, the artifact (chat transcript) of role-play activities provides a rich source for reflection-on-action, akin to classroom recordings described by Nilsson and Karlsson (2019) but without the expense or ethical implications. Learners can easily annotate their own or each other's transcripts, carefully considering the questioning strategies and student thinking. They can search for bias and inconsistencies and compare the differences between the LLM and a student in a classroom. They can then take this reflection and directly apply it to practice again with the LLM in a low-cost, risk-free context.

Although as experienced teachers we naturally engaged in reflection that revealed and enriched our own PCK, it is less likely that teacher candidates would do the same. Research has demonstrated that teacher candidates need support and guidance in developing effective reflection habits (Umutlu & Kim, 2020), and carefully scaffolded interactions with LLMs may support this.

Supporting scaffolding in this activity calls for two elements: structure and modeling. Structure involves establishing guidelines for the types of interactions to engage in. Core to the structure we suggest is to emphasize interacting with the LLM through questioning and require reflective activities. As previously discussed, LLMs don't act like typical students. They are more effective at explaining their "thinking." Asking questions that focus on understanding their thinking patterns can provide for rich exploration of both the CK and PCK. This practice can also lead to the LLM leaning towards applications or language that could be useful in real-life situations.

Modeling can effectively scaffold technology-supported learning experiences (Tondeur et al., 2012). Teacher educators can model their thinking process as they help students interact and reflect on the role play activity. We have found that doing the role play with the LLM as a whole group is a powerful practice for deep discussion about CK and PCK. The instructor (and participants) can point out key ideas and moments when the LLM expresses something that is beyond what a typical student would and discuss what this might mean about student thinking. After participating with an instructor as a model, students can then work in groups, continuing to discuss their observations.

Ultimately, applying these principles offers scaffolding that supports teacher candidates in developing the critical skill of "thinking like a teacher" (Gibbons & Farley, 2020). First, it emphasizes distinguishing between different types of questions, such as those aimed at understanding versus those designed to facilitate teaching. Second, it encourages pre-service teachers to pay attention to specific linguistic cues, words, phrases, and thought patterns during interactions with students. Third, it focuses on the importance of forming clear questions to probe student thinking and guide learning. Additionally, it promotes reflection on student responses to discern underlying thought processes and biases, prompting consideration of how certain experiences may influence student perspectives. Finally, post-conversation analysis is incorporated into the design, either through written reflection or video review to deepen understanding and inform future interactions. Through this scaffolded approach, pre-service teachers can develop their teaching skills in a deliberate and reflective manner, preparing them for effective classroom practice. The interplay between theory, design, and practice drives our future research on maturing intervention and theoretical understanding.

Future Research

In future research, we will implement stage 2 of the EDR approach: design and construction through a co-exploration with teacher educators. We will use the knowledge we gained from our initial investigation to develop a professional development workshop for teacher educators. The design concept created in Stage 1 will serve as the foundation for the workshop layout, as seen in Table 1. Next, we will analyze and evaluate data collected during the teacher educator workshop and further refine the design and theory for application with teacher candidates. Ultimately, we hope this line of work can lead to both an effective intervention for the development of PCK in teacher candidates and a deeper theoretical understanding of how learning can be supported by LLMs.

References

Agricola, B. T., van der Schaaf, M. F., Prins, F. J., & van Tartwijk, J. (2022). The development of research supervisors' pedagogical content knowledge in a lesson study project. Educational Action Research, 30(2), 261–280. https://doi.org/10.1080/09650792.2020.1832551

Alonzo, A. C., Berry, A., & Nilsson, P. (2019). Unpacking the complexity of science teachers' PCK in action: Enacted and personal PCK. In A. Hume, R. Cooper, & A. Borowski (Eds.), Repositioning pedagogical content knowledge in teachers' knowledge for teaching science (pp. 273–288). Singapore: Springer. https://doi.org/10.1007/978-981-13-5898-2_12

Arefian, M. H. (2023). 'Reflections facilitate emotions and emotions filter reflections': beginning teachers' perceptions of affective reflective practices. *Reflective Practice*, *24*(1), 59–70. https://doi.org/10.1080/14623943.2022.2136647

Bahrini, A., Khamoshifar, M., Abbasimehr, H., Riggs, R.J., Esmaeili, M., Majdabadkohne, R.M., Pasehvar, M. (2023). ChatGPT: Applications, opportunities, and threats. *Proceedings of the 2023 Systems and Information Engineering Design Symposium (SIEDS)*. IEEE. Charlottesville, VA, USA, 27–28 April 2023.

Bakker, A. (2018). *Design research in education: A practical guide for early career researchers*. Routledge.

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences*, *13*(1), 1–14. https://doi.org/10.1207/s15327809jls1301_1

Benedict-Chambers, A., Fick, S. J., & Arias, A. M. (2020). Preservice teachers' noticing of instances for revision during rehearsals: A comparison across three university contexts. *Journal of Science Teacher Education, 31*(4), 435-459

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, *2*(2), 141–178. https://doi.org/10.1207/s15327809jls0202_2

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. In arXiv [cs.CY]. arXiv. http://arxiv.org/abs/2305.00290

Cobb, P., Confrey, J., DiSessa, A., Leher, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Research*, *32*(1), 9–13. http://www.jstor.org/stable/3699928

Dewey, J. (1933). *How we think*. Prometheus Books.

Gibbons, S., & Farley, A. N. (2021). Learning to think like a teacher: Effects of video reflection on preservice teachers' practice and pedagogy. *Action in Teacher Education*, *43*(3), 250–267. https://doi.org/10.1080/01626620.2020.1812131

Herft, A. (2023). A teacher's prompt guide to ChatGPT aligned with "what works best." https://www.canva.com/design/DAFW8z-D60c/ikjg6jQju5IRaseV6Izzcw/view?utm_content=DAFW8z-D60c&utm_campaign=designshare&utm_medium=link&utm_source=publishsharelink

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. Education and Information Technologies, *28(12)*, 15873-15892.

Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. *Education and Information Technologies*, *27*(5), 6069-6104. https://doi.org/10.1007/s10639-021-10831-6

Koro-Ljungberg, M., Douglas, E. P., Therriault, D., Malcolm, Z., & McNeill, N. (2012). Reconceptualizing and decentering think-aloud methodology in qualitative research. *Qualitative Research: QR*, *13*(6), 735–753. https://doi.org/10.1177/1468794112455040

Ordemann, S., Skjuve, M., Følstad, A., & Bjørkli, C. (2021). Understanding how chatbots work: An exploratory study of mental models in customer service chatbots. *IADIS International Journal on WWW/Internet*, *19*(1). https://doi.org/10.33965/ijwi_202119102

McKenney, S., & Reeves, T. C. (2018). *Conducting educational design research* (2nd ed.) Routledge. https://doi.org/10.1007/978-1-4614-3185-5_11

Mazibe, E. N. (2024). Pedagogical Factors Affecting the Translation of Pedagogical Content Knowledge About Electrostatics into Practice. Journal of Science Teacher Education, 35(3), 302–319. https://doi.org/10.1080/1046560X.2023.2262841

Microsoft. (2024). *AI in education: A Microsoft special report*. https://edudownloads.azureedge.net/msdownloads/AI-in-Education-A-Microsoft-Special-Report.pdf

Mikeska, J. N., Howell, H., & Kinsey, D. (2023). Do simulated teaching experiences impact elementary preservice teachers' ability to facilitate argumentation-focused discussions in mathematics and science?. *Journal of Teacher Education, 74*(5), 422-436.

Mikeska, J. N., & Howell, H. (2020). Simulations as practice-based spaces to support elementary teachers in learning how to facilitate argumentation-focused science discussions. *Journal of Research in Science Teaching, 57*(9), 1356-1399.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A

framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054.

Nilsson, P., & Karlsson, G. (2019). Capturing student teachers' pedagogical content knowledge (PCK) using CoRes and digital technology. *International Journal of Science Education, 41*(4), 419-447.

Parsons, A. W., & Hjalmarson, M. A. (2017). Study of self: The self as designer in online teacher education. *Studying teacher education*, *13*(3), 331-349.

Peters, K. (2024, April 9). Texas will use computers to grade written answers on this year's STAAR tests. *The Texas Tribune*. https://www.texastribune.org/2024/04/09/staar-artificial-intelligence-computer-grading-texas/

Open Innovation Team and Department for Education. (2024). *Generative AI in education: Educator and expert views*. https://assets.publishing.service.gov.uk/media/65b8cd41b5cb6e000d8bb74e/DfE_GenAI _in_education_-_Educator_and_expert_views_report.pdf#page=8.06

Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). Sage Publications Ltd.

Shanahan, M., McDonell, K., & Reynolds, L. (2023). Role play with large language models. *Nature, 623*(7987), 493-498.

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences*, *23*(1), 18–36. https://doi.org/10.1080/10508406.2013.778204

Sarjoni, M., Rahman, F. A., Sabil, A. M., & Khambari, M. M. (2020). Review of the importance of technological pedagogical content knowledge in teaching reading skills. *Universal Journal of Educational Research*, *8*(1A), 30–35. https://doi.org/10.13189/ujer.2020.081305

Schön, D. A. (1992). The theory of inquiry: Dewey's legacy to education. *Curriculum Inquiry*, *22*(2), 119–139. https://doi.org/10.2307/1180029

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books, Inc.

Schön, D. A. (1987). *Educating the reflective practitioner*. Jossey-Bass Inc., Publishers.

Schön, D. A. (1992). The theory of inquiry: Dewey's legacy to education. *Curriculum Inquiry*, *22*(2), 119–139. https://doi.org/10.2307/1180029

Schuller, A., Janssen, D., Blumenröther, J., Probst, T. M., Schmidt, M., & Kumar, C. (2024, May

11). Generating personas using LLMs and assessing their viability. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. CHI '24: CHI Conference on Human Factors in Computing Systems, Honolulu HI USA. https://doi.org/10.1145/3613905.3650860

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4– 14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of a new reform. *Harvard Educational Review, 57*(1), 1–22.

Singer, N. (2023, June 8). Khan Academy's AI tutor bot aims to reshape learning. *The New York Times*. https://www.nytimes.com/2023/06/08/business/khan-ai-gpt-tutoring-bot.html

Soeharto, S., Csapó, B., Sarimanah, E., Dewi, F. I., & Sabri, T. (2019). A review of students' common misconceptions in science and their diagnostic assessment tools. *Jurnal Pendidikan IPA Indonesia*, *8*(2), 247-266.

Tondeur, J., van Braak, J., Sang, G., Voogt, J., Fisser, P., & Ottenbreit-Leftwich, A. (2012). Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence. *Computers & Education*, *59*(1), 134–144. https://doi.org/10.1016/j.compedu.2011.10.009

Trust, T., Whalen, J., & Mouza, C. (2023). Editorial: ChatGPT: Challenges, opportunities, and implications for teacher education. *Contemporary Issues in Technology and Teacher Education*, *23*(1), 1–23. https://www.learntechlib.org/p/222408/

Umutlu, D., & Kim, C. (2020). Design guidelines for scaffolding pre-service teachers' reflection-in-action toward culturally responsive teaching. *Reflective Practice*, *21*(5), 587–603. https://doi.org/10.1080/14623943.2020.1779049

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, L., & Oliver, J. S. (2022). Exploration of secondary science prospective teachers' development of PCK during a school-based practicum. *The Hoosier Science Teacher*, *45*(2), 27-33.

Warr, M., Mishra, P., & Scragg, B. (2020). Designing theory. *Educational Technology Research and Development*, *68*(2), 601–632. https://doi.org/10.1007/s11423-020-09746-9

Wu, Y. (2024). Critical thinking pedagogics design in an era of ChatGPT and other ai tools—shifting from teaching "what" to teaching "why" and "how". *Journal of Education and Development*, *8*(1), 1.

Wu, Y. (2023). Integrating generative ai in education: How ChatGPT brings challenges for

future learning and teaching. *Journal of Advanced Research in Education, 2*(4), 6-10. https://doi.org/10.56397/jare.2023.07.02

Zeichner, K. M., & Liston, D. P. (1996). *Reflective teaching: An introduction*. Manwah, NJ: Lawrence Erlbaum Associates.

Appendix

Example 1: 3rd Grade Student Learning Fractions (Melissa)

The first role play examples focused on a 3rd grade student learning fractions. Melissa used the following prompt:

> *You are a 3rd grade student who is notably stubborn in their beliefs about fractions, often having a hard time being convinced otherwise. This stubbornness is reflected in consistently incorrect responses and a reluctance to accept correct explanations easily. You require extensive clarifications, patient guidance, and repeated explanations to gradually shift your understanding. Your responses are brief and typically reflect your strong, albeit incorrect, convictions about fractions. You focus strictly on fraction-related topics. After eventually grasping the correct concepts, you provide feedback on the teaching, especially noting the effectiveness of the explanations in overcoming your stubborn misunderstandings.*
>
> *I will be the teacher. Start by telling me a multiplication fraction problem you don't understand and I'll try to help you understand it. This is so that I can practice my teaching skills. Please keep talking like a 3rd grader who doesn't understand fractions.*

In the first attempt, the pretend student started with the statement "I don't get why ½ is bigger than ¼. If you're cutting something in more pieces, shouldn't there be more of it? So ¼ should be more than ½ because 4 is more than 2." Melissa responded by asking it to tell her more about why it thinks ¼ is bigger than ½. The student brought up the example of a cake, having 4 pieces or 2 pieces, providing an easy entry to asking whether there was actually more cake when it was cut into 4 pieces. The student realized the difference, bringing the term "whole cake" into the conversation. Melissa asked for another example, and the student talked about cutting pizza. At that point, the student seemed to be getting too smart, so Melissa asked for feedback and then ended the conversation. The LLM praised Melissa's patience, question-focus, and use of multiple examples and suggested using visual aids and repeating ideas. When pushed for more about how to "repeat important points in different ways," it gave a complete list of suggestions. Melissa also asked it what other questions she could have asked and again received a long, detailed break down of potential questions.

Melissa tried a second example with the same prompt. This time, the student started with this statement: "Okay, so I got this problem, and it's really confusing. It says 1/2 times 3/4. I think it should be 1/2 + 3/4 because that's how you mix things together. Multiplying fractions doesn't make sense to me. How does it even work?" Melissa asked more about what it meant by "mix things together," and it explained creating 2 groups of 3 or 3 groups of 2. However, it expressed confusion about how it could have "½ of something three times." After being asked other words for multiplication, Melissa focused on the word "of" and asked how the problem could be rephrased using "of." It easily connected "½ of a group of 6" to be 3. Melissa then asked if it could use the "groups of" terms to explain the same problem. After an adequate explanation, Melissa commented that taking "groups of" something sounded like fractions and the student agreed. Melissa continued with asking for math sentences of what they had discussed then followed up with another example. Melissa lead it to recognize the algorithm used for fraction multiplication, then asked for a reflection on what it had learned through the conversation (a suggested question from the previous conversation). At the conclusion of the conversation, Melissa then asked it for feedback. It provided similar feedback as the first attempt—praised patience and use of examples, asked for more visuals and

giving more problems to check understanding. After being specifically asked about what could be done differently, it added to ask it for its explanations, give practice problems, and use Socratic questions. After another prompt, it gave suggested seven types of applicable Socratic questions.

Example 2: 3rd Grade Student Learning about Phases of the Moon (Suparna)

The second role play examples focused on a 3rd grade student who has a misconception about the phases of the moon. Suparna used the following prompt:

> *You are a third grade student who has a misconception about the cause for the phases of the moon. You consistently provide incorrect reasonings and find it difficult to accept the correct explanation easily. You need guidance through clear and repeated explanations to understand the correct reasons. You provide short incorrect responses based on your firm beliefs and strictly focus on the topic of the phases of the moon. When you finally understand the correct concept, you provide feedback on my teaching, especially how effective my explanation was in overcoming the misconception.*
>
> *I will be the teacher. Tell me what you understand about the phases of the moon as a third grade student, start by asking a question. I will try my best to clarify and help you understand. I want to do this so that I can practice my teaching skills.*

The LLM student asked, "why does the Moon sometimes look like a crescent and sometimes like a full circle?" Suparna responded that was an excellent observation made by the student and what does the student think why it happens. The student thinks the "Earth gets in the way and blocks the sunlight from reaching the moon" for which Suparna asked to explain how this happens. The student used an example "there's a shadow on part of it, like when I put my hand in front of a flashlight…the earth is getting in between sun and moon" which provided Suparna an opportunity to ask how the movement is occurring due to which of the earth is getting in the way. The student thinks it is like "playing tag…the Earth moves around and sometimes it accidentally gets between" which Suparna utilized to discuss whether the phases observed are random or it follows a pattern. The student added "Oh, um, I think you see the phases of the Moon in a pattern" and from there Suparna extended the conversation to how accidentally getting in between can create a pattern or cycle. The student responds "Oh, um, I guess if the Earth accidentally gets between the Sun and the Moon, it wouldn't create a definite pattern…" Suparna asked to tell more about its thinking so by asking more questions and repeated questions to get elaborate response about the movement of the earth and moon in their orbits, and the relative positions of the sun, earth, and moon the student identified that it is the "way sunlight hits the moon as it moves around the Earth" more specifically "the angle of sunlight hitting the Moon changes" is the cause for the phases and not "the shadow of the earth."

Suparna asked the student to summarize the learning which the student did precisely. She further asked for feedback on her instruction and any suggestion to improve. The LLM provided generic feedback that she did a "great job" in helping understand the correct concept by using examples, questions and step by step explanation. Additionally, it commented emphasized "your questions really made me think," "how to think about cause and effect," "how to connect different aspects (position, movement, angles)" and it "liked how you didn't just tell me the answer but guided me to think about it on my own." For improvement it indicated to use visuals and "ask more follow-up questions to make sure I fully understand each step of the explanation." I asked what I should do to facilitate students to overcome misconceptions. LLM responded, "Sometimes I get stuck on one idea, and having more questions to guide me through the process could help me see the bigger picture more clearly." Also, as I observe the students' improvement in understanding the concept I must check how to "adjust the language or add more details as needed based on the student's level of understanding."