Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mathematics Teacher Interaction Styles with Generative Artificial Intelligence: Personalization in Word- Problem Design

Theodora Beauchamp

theob@smu.edu

Southern Methodist University

Candace Walkington

Southern Methodist University

Tiffini Pruitt-Britton

American Institutes for Research

Research Article

Keywords: Generative AI (GenAI), epistemic agency, context personalization, generative authority

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11210562/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

As generative artificial intelligence (GenAI) plays an increasing role in K-12 mathematics classrooms, understanding how teachers collaborate with these technologies during instructional design is essential. Drawing on frameworks related to personalization theory, Technological Pedagogical Content Knowledge, and adaptive epistemic agency, this study examined how seven seventh-grade mathematics teachers interacted with ChatGPT to personalize mathematics word problems to students' interests. Analyses of teacher-GenAI interactions identified three interaction styles: AI-Dominated-Passive Prompt Personalization, Teacher-Dominated-Rigid Personalization, and Human-AI partnership-Collaborative Personalization. Students' ratings of the resulting personalized problems varied across teachers and interaction styles, with the strongest positive patterns occurring among teachers characterized by Collaborative Personalization. Collaborative personalization illustrates how teachers may maintain epistemic authority for evaluating mathematical, pedagogical, and contextual appropriateness while selectively granting GenAI generative authority to expand personalization possibilities. These findings highlight the importance of preparing teachers to critically evaluate and collaboratively engage with GenAI during instructional design.

1. Introduction

Generative artificial intelligence (GenAI) is reshaping mathematics instruction by helping teachers quickly create and personalize mathematical tasks. Tools like ChatGPT allow for tailoring of mathematics content to students' interests, cultural context, and academic needs (Seo et al., 2024; Uğras, 2024). GenAI's value depends on how teachers guide the AI's processes of personalization and how they engage in joint decision-making (Biton & Segal, 2025; Drushlyak et al., 2025). However, prior research has largely emphasized what GenAI can generate rather than how teachers interact with these tools during task design. For mathematics word problems in particular, contextual personalization requires maintaining mathematical structure while adapting narrative elements to be coherent, age-appropriate, and relevant. This creates a central tension in teacher-GenAI collaboration: teachers must decide what to delegate to the AI model, what to constrain, and how to evaluate and revise outputs when they are unrealistic or misaligned conceptually.

The increasing capacity of GenAI to rapidly generate personalized content makes teachers' professional judgment especially important. Personalization is not simply the insertion of an interest, name, or popular reference into an existing mathematics problem. Teachers must determine whether a generated context is authentic and recognizable to students while also ensuring that quantities, relationships, and mathematical demands remain plausible within the new context. GenAI may provide a broader range of interest-based contexts for teachers to integrate into their instruction, particularly when students' interests fall outside teachers' own knowledge or experiences, but these possibilities still require evaluation. Effective personalization involves both generative and evaluative work: generating possibilities that may connect with students while determining whether those possibilities are mathematically, pedagogically, and contextually appropriate.

These decisions also raise questions about how authority is distributed during teacher-GenAI collaboration. Teachers make many decisions during the interaction process, such as when to delegate to AI, what information to provide, when to accept an output, when to question it, asking AI for multiple responses, and when to override AI completely. Teachers may tightly constrain GenAI by specifying most features of a personalized problem, delegate substantial responsibility for generating contexts and references to the model, or move between direct and delegating as they evaluate successive outputs. Just as important, greater teacher control does not necessarily indicate more productive collaboration, just as greater reliance on GenAI does not necessarily indicate an absence of teacher expertise. What may be more important is how teachers adaptively determine when to exercise professional judgment and when GenAI's generative capabilities can extend the possibilities available to them. Examining these moment-by-moment decisions can provide insight into how teachers maintain epistemic responsibility for mathematics task design while selectively sharing generative authority with GenAI.

The present study examines these processes by investigating how 7 seventh-grade mathematics teachers interacted with ChatGPT while personalizing mathematics word problems to students' interests. We examine teachers' prompting and decision-making practices, students' perceptions of the resulting personalized problems, and differences in how teachers distributed responsibility between themselves and GenAI during the personalization process. Our study seeks to better understand what teacher-GenAI collaboration looks like during authentic mathematics task design.

2. Theoretical Framework

The present study draws on two complementary theoretical perspectives: technological pedagogical content knowledge and its AI-related extensions, which provide a lens for considering the knowledge teachers bring to GenAI-supported instructional design. Secondly, epistemic agency provides a lens for examining how teachers exercise judgment and decision-making authority during interaction with GenAI.

2.1 TPACK, AI-TPACK, I-TPACK

The TPACK framework (Mishra & Koehler, 2006; 2026) emphasizes the intersection of technological, pedagogical, and content knowledge in effective teaching. TPACK highlights the complex knowledge base required to teach effectively with technology, including not only digital fluency (Technological Knowledge, TK), but also instructional design (Pedagogical Knowledge, PK) and mathematical expertise (Content Knowledge, CK).

Recent frameworks, such as AI-TPACK (Ning et al. 2024), have expanded TPACK to include considerations specific to AI. Artificial Intelligence Technological Pedagogical Content Knowledge (AI-TPACK) by Ning et al. (2024) extends the TPACK framework from Mishra and Koeler (2006) by explicitly incorporating AI intelligence (algorithmic reasoning, generative capability, adaptivity, and multimodal output) as a new component that teachers must understand and orchestrate. Wu et al. (2026) developed a mathematics-specific AI-TPACK measure for preservice teachers engaged in mathematics problem

posing. Teachers' confidence in using GenAI was connected not simply to technical skill but how they learned to use GenAI pedagogically (Wu et al., 2026; Mosia et al., 2026). Collectively, these studies suggest that knowing how to operate or prompt an AI system does not necessarily indicate readiness to use it instructionally. Rather, teachers must be able to connect TK with mathematical content and pedagogical reasoning and critically evaluate when and how AI-generated materials should be incorporated into instruction. This distinction is supported by research with in-service teachers showing that technological pedagogical knowledge (TPK) may serve as an important link between technological proficiency and broader TPACK to GenAI competence (Aga, Sawyer, & Wolfe (2024); Blundell et al. (2026); Cindilli et al. (2026); Hussein, Daher, & Assali (2026); Xie & Luo (2026)).

Additional literature important for teacher preparation and professional development is from Dogan (2026), which extends I-TPACK into a professional development framework that integrates AI-specific technological, pedagogical, and content knowledge with ethical considerations and emphasizes active learning, modeling, coaching, feedback, and reflection. Empirical findings provide some support for this approach. Hussein et al. (2026), for example, found that TK appeared to provide an entry point into GenAI use, while pedagogical integration, critical evaluation, and assessment-related competencies developed more gradually.

2.2 Epistemic Agency

As teachers engage with AI, they also share epistemic agency with the AI (Wu et al., 2025). Wu et al. (2025) called this a "symbiotic learning environment" (p.366), design in which human-AI partnerships are formed. Teachers' interaction styles can be understood as varying expressions of epistemic control and shared agency. Within this shared epistemic environment, Wu et al. (2025) shared that the ways individuals interact with GenAI may be shaped by their epistemic stance, or the position they adopt regarding how knowledge is acquired and validated (p.360). Taken from Kuhn et al. (2000), Wu explained that learners can take an absolutist stance, where they unreflectively accept generative AI output as truth and as the authority. They can take a multiplist stance, where they engage in deeper inquiry relating to GenAI's outputs, attempting to understand the reasons behind and applicability of the GenAI's output. And finally, they can take an evaluativist stance, where they consider and integrate multiple ideas given by AI and peers, critically evaluating each proposed output.

Recent empirical research further suggests that teachers' judgments about AI are situated within particular instructional contexts rather than governed by fixed rules. Karumbaiah et al. (2026) examined how teachers reasoned about fairness and potential harms associated with education AI and found that teachers engaged in conditional, context-sensitive reasoning, drawing on their knowledge of students, classroom dynamics, and instructional goals when evaluating AI-supported practices. Rather than reaching a uniform judgement about whether particular uses of AI were acceptable, teachers considered the conditions under which benefits, risk and potential harms might warrant different decisions. Importantly, expressions of uncertainty did not necessarily indicate an absence of professional judgment; instead, teachers used uncertainty and deliberation to identify what additional information or conditions were needed before reaching a decision.

The importance of maintaining human judgment during Gen-AI-supported activity is also reflected in another recent scholarship on epistemic authority. Özer el al. (2026) argues that the central concern surrounding GenAI is not simply whether AI contributes to the production of an output, but the extent to which humans also delegate responsibility for evaluating and judging that output. When AI becomes both a producer and evaluator, epistemic authority may shift away from the human agent, weakening human responsibility and accountability.

3. Literature Review

3.1 Personalization in Mathematics Education

Personalization can be defined as adapting instruction to connect academic content with students' individual interests and experiences (Author, Date). In mathematics, interest-based personalization may involve embedding mathematical relationships within contexts connected to students' hobbies, activities, popular culture, or other areas of their lives. Such connections can help make mathematics more personally meaningful and highlight its relevance beyond the classroom (Bernacki & Walkington, 2018; 2026; Author, Date). Some possible limitations are also important to note. Creating different versions of word problems may be time-consuming for teachers (Author, Date). Recent conceptualizations of personalized learning emphasize that personalization involves more than modifying instructional materials for individual learners. Indeed, Bernacki and Walkington (2026) propose an extensible model of personalized learning in which information about the learning environment is informed by specified learning outcomes. One application of this framework is context personalization, in which instructional tasks are adapted to reflect students' everyday interests. Bernacki and Walkington (2026) describe depth, grain size, and ownership as important design considerations: personalization can vary in how authentically it represents students' engagement with an interest, how specifically it reflects an individual interest, and how much control students have over the personalized connection.

Emerging research suggests that these considerations remain important when GenAI is used to personalize mathematics tasks. Author (Date) found that GenAI supported teachers in efficiently generating interest-based mathematics tasks, but AI-generated contexts varied in accuracy and authenticity. Teachers needed to evaluate and refine generated problems to ensure that personalization reflected students' interests and experiences rather than relying on surface-level connections.

The emergence of GenAI may address some of the practical barriers associated with creating personalized instructional materials by allowing individual problem contexts to be generated more efficiently. Tasdelen and Bodemer (2025) investigated AI-generated, context-personalized learning materials and mathematics tasks in primary school classrooms. In a 2 x 2 mixed design with 114 students in Germany, the researchers compared context-personalized and standard learning materials and tasks. The personalized condition used AI-generated materials tailored to students' individual interests, whereas the standard condition used AI-generated materials based on more typical textbook

contexts. Students who received context-personalized learning materials and tasks demonstrated greater intrinsic motivation, interest, and learning performance than students receiving standard materials and tasks. These findings provide emerging evidence that GenAI may make interest-based context personalization more scalable while retaining some of the motivational and learning benefits associated with personalization (Tasdelen & Bodemer, 2025).

Although GenAI may reduce the time required to generate individualized mathematics tasks, the availability of generative tools does not ensure that the resulting personalization will meaningfully reflect students' interests. The depth and specificity of personalized contexts may depend partly on how teachers communicate learning information to GenAI, evaluate generated contexts, and revise those contexts during the personalization process. GenAI introduces not only a mechanism for scaling context personalization but also a new instructional design process in which teachers make decisions about when and how AI-generated adaptations appropriately represent students' interests while preserving the mathematical goals of the original task.

3.2 Teacher attitudes toward GenAI

How teachers position AI within instructional design may provide insight into teachers' attitudes toward GenAI. Wang et al. (2026) systematic review of 28 empirical studies published between 2015 and 2025 examined how teachers collaborate with AI, including GenAI, to design learning tasks. The review provided evidence of teachers' relationship with AI during the design process, highlighting that GenAI was most commonly used for lesson planning, generating prompts/content, and creative ideation. Wang et al. (2026) found that AI was used mostly to assist (33.3%) and not as a co-designing or dialogic partner (13.3%). Existing research does not adequately explain how teachers decide when to accept, review, or reject AI-generated outputs; these choices are epistemological rather than merely procedural decisions, and professional development needs to move beyond teaching educators how to operate AI. Teachers need opportunities to develop design thinking, critical data interpretation, ethical judgement, and the ability to question and refine AI-generated ideas when deemed necessary (Wang et al., 2026).

Similar conclusions emerge from O'Ceallaigh and Murphy's (2026) systematic review of GenAI-specific TPK in teacher education. Their findings suggest that effective GenAI integration requires knowledge extending beyond basic technical competence, including pedagogical reasoning, critical evaluation, ethical awareness, and the ability to design instruction in which AI use supports rather than replaces pedagogical goals. Consistent with Wang et al. (2026), O'Ceallaigh and Murphy caution against both uncritical reliance on GenAI and rejection of the technology. Instead, teachers need capacity to critically interrogate AI-generated content and make intention pedagogical decisions regarding when and how it should inform instruction. These studies, together, suggest that teachers' relationship with GenAI cannot be understood solely through attitudes toward technology or their ability to operate it. Rather, meaningful GenAI integration requires teachers to exercise professional judgment about pedagogical values, limitations, and appropriate role of AI-generated contributions.

3.3 Research Purpose and Research Questions

Emerging research demonstrates the potential of GenAI to support personalized learning and mathematics task design, while frameworks such as TPACK and I-TPACK identify the knowledge teachers need to integrate AI effectively into instruction. Research on human epistemic agency further emphasizes the importance of critically evaluating AI-generated content (Özer et al., 2026) and adaptively evaluating when to rely on, question, or revise AI contributions (Wu et al., 2025). However, less is known about how teachers enact these forms of agency during authentic instructional design in mathematics education.

This present study advances this work by examining how epistemic authority is negotiated through teachers' moment-to-moment interactions with GenAI for mathematical problem-posing, and by distinguishing teachers' responsibility for evaluating instructional decisions from the generative authority selectively granted to GenAI. Our research questions are: (1) What styles of interaction do teachers adopt when co-designing personalized mathematics story problems with ChatGPT? (2) How do these styles relate to the students' interest in and perceived closeness of the generated problems to their interests? (3) How do teachers exercise epistemic agency when deciding when to retain decision-making authority and when to delegate generative authority to GenAI during mathematics problem personalization?

4. Method

4.1 Study Design

This study employed a mixed-methods multiple case study design. Case study research is appropriate for examining a bounded phenomenon in depth and for considering patterns both within and across cases (Creswell & Poth, 2018). Each of the seven teachers was treated as an individual case, allowing teacher-GenAI interactions to be examined within each teacher's personalization sessions and subsequently compared across cases. The qualitative component focused on teachers' observable prompting and revision practices and their articulated reasoning during interactions with ChatGPT. Qualitative data were analyzed using thematic analysis following Braun and Clarke (2006) through an iterative process of coding, categorizing, and identifying patterns across the data (Creswell & Poth, 2018). This allowed the analysis to capture not only what the teachers asked GenAI to do, but also the reasoning underlying their decision to delegate, constrain, accept, or revise its output. The quantitative component examined students' rating of the personalized problem on a 1–5 scale as they solved the problem for homework.

4.2 Participants

The study included seven 7th -grade mathematics teachers working across five school districts in multiple regions of the United States, including the South, West, Northeast, and Midwest. We recruited five teachers through an email sent to 7th-grade teachers on a mailing list of educators currently using ASSISTments. The recruitment email directed teachers to the study website that described the project and included a Qualtrics screening survey. The survey collected information about teachers' schools,

students, and instructional contexts. Seventeen teachers completed the screening survey. We invited teachers who taught more than 50 seventh-grade students to continue in the recruitment process, resulting in 11 eligible teachers. Of these, 5 of the 11 chose to complete the remaining recruitment requirements, including obtaining a letter of support from their school principal, and enrolled in the study. Two additional 7th grade mathematics teachers were recruited from a middle school with which one of the researchers had previously worked. These teachers had not previously used ASSISTments but expressed interest in incorporating the platform into their instruction. The final sample consisted of seven teachers: five identified as female and two as male. Six identified as White and one identified as multiracial, Black and White. Five teachers reported holding bachelor's degrees in education, one held a bachelor's degree in mathematics, and one reported holding both a bachelor's and a master's degree. Teachers provided informed consent and received compensation for participating in the study. The university's Institutional Review Board approved the study.

Teachers are represented in Table A1 in Appendix A, which summarizes participants' teaching experience and prior experience with GenAI. The seven participating teachers' teaching experience ranged from 3 to 22 years (M = 12.0 years). Despite their varied levels of classroom experience, most participants entered the study with limited or no prior experience using GenAI for instructional purposes (Appendix A). Teachers B and D reported essentially no prior use, while Teachers A, E, and F characterized their previous experience as minimal. Teachers C and G reported somewhat greater familiarity with GenAI; however, their experiences differed in purpose. Teacher C had previously used GenAI to generate supplemental mathematics problems and problem sets, whereas Teacher G primarily explored ChatGPT to understand its capabilities and monitor how students might use it. Although participants brought substantial variation in years of teaching experience, sustained experience using GenAI to personalize mathematics problems was limited across the sample. This variation provided an opportunity to examine how teachers with different levels of professional and GenAI experience interact with the technology during the personalization process.

4.3 Measures and Tasks

The teachers participated in two ChatGPT 4.0-supported sessions to generate personalized word problems based on practice problems that were already in the 7th -grade *Illustrative Mathematics* curriculum, from Units 4 (Proportional Relationships and Percentages) and 6 (Expressions, Equations, and Inequalities). Personalization was conceptualized as modifying contextual features of a problem to create connections with students' interests and experiences while maintaining the math students were expected to solve. Teachers were encouraged to personalize problem-based student characteristics such as age, geographic location, and interests in popular culture, including sports, music, food, video games, and other topics they believed would be meaningful to their students.

Before the personalization sessions, teachers participated in professional development that included strategies for identifying students' interests for use in personalized instruction. Teachers learned multiple approaches to understanding their students' interests, including interest surveys, interviews, and informal conversations with students. However, teachers were not required to use a standardized

interest-assessment procedure. Consistent with the study's emphasis on teachers' professional knowledge of their students, teachers determined how and when to gather information about students' interests and drew upon both information collected from students and their existing knowledge of their students when personalizing mathematics problems. Consequently, the source and specificity of student-interest information available during the personalization sessions varied across teachers.

Each teacher participated in two 120-minute personalization sessions using ChatGPT-4 (Table A2 in Appendix A). The number of problems personalized by teacher varied, as the time allotment included interactions with AI, which varied across problems and teachers. Each teacher posed an average of 20.3 problems (*SD* = 8.1) during the two sessions, for a total of 144 personalized problems posed (Table A2). The personalized problems were later delivered to the teachers' 534 7th grade students in the ASSISTments online homework platform (Heffernan & Heffernan, 2014). In ASSISTments, students in the teachers' classes were given a choice between three different topics that their assignment could be personalized to (e.g., sports, after-school activities, video games). Once the student made a choice for the assignment, that choice persisted for all the problems in that assignment. The broad topics were chosen by the teachers.

4.4 Coding of Individual Teacher Moves

We took all teacher prompt logs with ChatGPT, as well as their transcribed verbal utterances (i.e., things that they said out loud) as they were working with ChatGPT to personalize mathematics problems. In order to analyze the logs and transcripts, they were aligned so that each unit of analysis consisted of a teacher's prompt to ChatGPT together with any verbal utterances associated with that prompt. Teachers were encouraged during the sessions to speak out loud to discuss their decision-making, and the researchers would also often proactively question them about why they were prompting ChatGPT in particular ways.

The data corpus had previously been coded with a scheme for individual teacher prompting moves (Table A3 in Appendix A) that was developed in another study with this same corpus (authors, date). That study's write-up can be referenced for more detailed information about the process of coding individual teacher prompts to the GenAI. This article's contribution is not the coding scheme in Table A3, as a prior study described and analyzed this coding scheme. Rather, the purpose of the present study is to use this scheme of specific teacher moves to generate *higher-level teacher interaction styles that capture epistemic agency and teacher-AI collaboration*, and then link these interaction styles to student ratings of personalized problems. Indeed, our prior work has looked at each of the many categories of teacher moves in Table A3 separately, without considering *how* and *in what combination* they are enacted, and what this signals about human-AI collaboration. In addition, we also did not previously examine *patterns of these interactions within teachers*, and we thus treated each instance of a prompting move as independent for another instance, although the research team noticed and discussed really significant and important differences between teachers' styles with GenAI interaction anecdotally while the study was going on. Thus, although we use the individual teacher prompting moves coding scheme

as an initial pass at understanding this data, it is not the primary focus of this study and it is only provided in Appendix A for reference.

4.5 Student Ratings of Personalized Problems

The teachers worked in schools serving diverse student populations. Across schools, the percentage of students from historically marginalized racial and ethnic groups ranged from 13% to 99% (M = 49%), while the percentage of students from socioeconomically disadvantaged backgrounds ranged from 3% to 55% (M = 30%). Collectively, the participating teachers taught approximately 534 seventh-grade students. We were able to collect 1–5 problem ratings from all 534 students. In addition, of these 534 students, 197 had an additional layer of parental consent allowing the research team to collect personal demographic information. Among these 197 students, 76 identified as female, 88 as male, 6 selected another gender identity or preferred not to respond, and gender data were missing for 27 students. For race and ethnicity, 121 students identified as White, 25 as Hispanic, 7 as Black, 6 as Asian, and 10 as another racial or ethnic category; data were missing for 28 students.

When students solved the personalized problems in ASSISTments, they responded to two questions after one-third of problems: (1) “On a scale of 1–5, how *interesting* did you find this last problem?” and (2) “On a scale of 1–5, how *closely* did this problem match things you are interested in?” This item was only given for approximately one-third of the problems because students would tire of constantly answering this question if they got it every time. The 1–5 ratings had a mean of 2.51 and a standard deviation of 1.26. These ratings were collected for a total of 53 personalized problems. We fitted mixed-effects regression models to predict the students’ 1–5 interestingness and closeness ratings (Raudenbush & Bryk, 2002). The unit from which each problem originated in the 7th grade *Illustrative Mathematics* curriculum was included as a fixed effect, while teacher identifier was included as a random intercept to account for problems being personalized by one of seven teachers. The resulting teacher-level random intercepts indicated whether problems personalized by a particular teacher tended to receive ratings above or below the overall mean for interestingness and closeness, after accounting for the unit. These numbers gave us a signal for how students tended to respond to each teacher’s attempts at writing personalized problems, to determine if there were teachers who wrote more versus less effective personalized problems in terms of student ratings. These random intercepts tended to range from negative 0.2 to positive 0.2, on a 5-point rating scale. This corresponds to plus or minus one-sixth of a standard deviation on that scale that could be related to the teacher’s overall problem-writing approach. Students’ actual performance on the problems would be another important signal, but variations were too numerous for this approach to be possible in the present study. The quantitative data should be interpreted as exploratory and hypothesis-generating only, given the small sample size and the relatively small teacher-level effect sizes.

4.6 Determining Interaction Style Themes

A qualitative thematic analysis following Braun and Clarke’s (2006) approach examined how teachers interacted with GenAI while personalizing mathematics problems and how patterns of interaction

differed across teachers. The ways in which each individual teacher had patterns and sequences of codes related to the prompt coding scheme established (Appendix Table 2) in the previous study (authors, date) was examined alongside teachers' post-session reflections, to identify broader regularities in how each teacher interacted with GenAI. Analysis focused not only on the frequency of particular prompting codes within teacher, but also on how coded moves occurred across the personalization process and how teachers described their own reasoning and decision-making following the writing sessions. Teacher cases were compared to identify patterns in how teachers directed personalization decisions, delegated the generation of possibilities to GenAI, evaluated GenAI-generated content, and then revised or redirected outputs. The post-writing session reflections were used to contextualize and help interpret these patterns rather than as a separate basis for assigning teachers to categories.

Through iterative within- and cross-case comparison of the coded interaction data and teachers' reflections, three higher-order patterns of teacher-GenAI interaction were identified: Teacher-Dominated-Rigid Personalization, AI-Dominated Passive Prompt Personalization, and Human-AI Partnership Collaborative Personalization. The interaction styles were not predetermined categories, nor were the teachers assigned to a style based on a single code or numerical threshold. Instead, the categories represented recurring configurations across teachers' coded interaction patterns in their account of the personalization process. Thus, the three interaction styles emerged as higher-order interpretations of patterns across the previously coded data, informed by teachers' reflections on their decision-making. These codes were determined and assigned to the 7 teachers by a single coder who had been part of the larger team that coded the individual teacher prompting moves. This coder's coding of the interaction styles by teacher were reviewed by the team's lead for clarity and appropriateness, to ensure that they were distinct enough and that the evidence being used to assign an interaction style to a teacher was appropriate.

5. Results

To understand teacher interaction styles with GenAI, we present findings across three sources of evidence: teachers' observed interactions with LLM during problem personalization, students' ratings of problem Interestingness and Closeness, and teachers' reflections during final interviews. The findings were organized into three conceptual categories, integrating both quantitative and qualitative results. To convey the core ideas shaping this study's findings, we developed a concept map (Fig. 1) that visualizes three teacher interactive styles when using GenAI: (1) AI-Dominated- Passive Prompt Personalization; (2) Teacher-Dominated- Rigid Personalization; and (3) Human-AI Partnership- Collaborative Personalization. These styles reflect different configurations of agency, instructional priorities, and integration of TPACK dimensions—particularly TK, PK, and CK.

5.1 Interaction Style 1: Teacher-Dominated Rigid Personalization

Teachers D and E demonstrated a Teacher-Dominated interaction style characterized by rigid personalization, in which the teacher maintained primary decision-making authority over how GenAI transformed the mathematics problems. Rather than allowing the AI to generate personalization outputs, these teachers frequently entered the interaction with predetermined ideas about what should be personalized and how exactly it should be personalized, and continuously constrained the GenAI accordingly. Their interactions reflected substantial teacher agency, but a limited sharing of that agency with the GenAI.

Teacher D frequently specified exactly what was to be personalized in a given problem, giving the GenAI little decision-making authority. For example, Teacher D directed GenAI to “change the name Zander to Jayden” and, in another revision, to “combine the first two sentences.” Teacher D’s individual prompting moves to the AI often focused on lower-level tasks like managing the readability of the personalized problems (21.2% of teacher moves) and having the AI include student names in the problem (15.15% of teacher moves). At the same time, this teacher demonstrated no coded attention in her prompting to have the AI manage the contextual realism of the story scenarios (0% of teacher moves) and also paid little attention in her prompting to managing the mathematical accuracy (3.03% of teacher moves). This is consistent with her stated instructional priority of keeping problems accessible and concise: “The lengthier the word problem, the less interested they get. They like them short and sweet.”

In Fig. 2, Teacher D’s interaction illustrates Rigid Personalization through her progressive specification of nearly every feature of the problem context. Rather than allowing GenAI to determine the personalization details, she repeatedly constrained the output by specifying the student name, brand, price, store, and wording, leaving GenAI to generate within an increasingly teacher-defined personalization space. Only the first prompt for the initial generation is something the teacher could not have easily done herself through simple editing of the problem, and she likely could have accomplished it with less time and trouble without using an LLM.

Teacher E similarly retained substantial control over personalization, but focused on a different set of criteria. Teacher E supplied specific personalization references and gave the AI particular numbers to use to ensure the problem was realistic in the real world. Nearly half of Teacher E’s coded moves (47.62%) involved concerns about unrealistic numbers, with 19.05% addressing unrealistic quantities (e.g., something being measured in a unit it would not reasonably be measured in). The teacher giving the AI specific popular culture references to include accounted for 23.81% of the coded moves. In contrast to Teacher D, there were no coded instances involving the teacher using prompts to have the AI adjust the problem based on student age, student location, typical student names, problem readability, or mathematical accuracy checking.

In Fig. 3, Teacher E demonstrated Rigid Personalization through repeated redirection of GenAI-generated contexts. In the interaction, she progressively narrowed or replaced GenAI’s suggestions by specifying the music genre, subgenre, date range, and artist, at times reversing earlier specifications until the generated context aligned with her preferred personalization and what she perceived to be most

realistic. A central problem she was experiencing was that the number of total songs on Spotify kept being represented by the AI as being unrealistically low. Rather than give the AI this feedback, she instead chose to give the AI a series of increasingly specific prompts where she attempts to fix the problem herself, essentially through direct editing of the problem text, similar to Teacher D. Also, Teacher E's repeated "nice try" wording is qualitative evidence because it shows her evaluative stance toward GenAI's proposal.

Some differences between Teacher D and E are evident in students' ratings of their personalized problems. Teacher D's random intercepts were close to the overall mean for both Interestingness (0.007) and Closeness (0.063). Teacher E, on the other hand, had substantially negative random intercepts for both Interestingness (-0.203) and Closeness (-0.191). Teacher dominance itself did not equally correspond with negative student ratings. The variation between Teachers D and E suggests that teacher dominance alone may not account for how students perceived the personalized problems, although it may create conditions where students are not rating problems particularly highly.

5.2 AI-Dominated: Passive Prompting

Teachers F and G demonstrated the AI-dominated interaction style characterized by passive prompt personalization, in which greater decision-making authority for generating personalization references was delegated to GenAI. In contrast to teacher-dominated personalization, Teachers F and G more frequently relied on the LLM to propose how and determine itself exactly how the original mathematics problems could be contextualized around students' interests.

Teacher G demonstrated a more passive form of AI-dominated personalization. Teacher G's coded teacher moves reflected attention to having the AI personalize the problem to students' locations (17.78% of teacher moves) and prompting to maintain contextual realism (15.56%). Different from Teacher F, Teacher G averaged 2.73 prompts per problem and 2.19 minutes per problem, the shortest average interaction time among the seven teachers. During personalization, he relied more heavily on GenAI to generate references itself with little input form him, and stated, "Just going to go generic with it. See what happens. See what turns up." Teacher G had strongly negative random intercepts for both interestingness (-0.200) and closeness (-0.182).

In Fig. 4, Teacher G demonstrated Passive Prompt Personalization through broad personalization requests that delegated substantial generative decision-making to GenAI. Rather than identifying specific personalization references, Teacher G frequently supplies only a general category, such as asking GenAI to make problems " about food" or "about animals," leaving the model to determine the specific context that would replace the original problem. In Fig. 4, we see Teacher G asking multiple times for the GenAI to brainstorm different foods, rather than intervening with a suggestion himself when he was dissatisfied with its results. Across these interactions, GenAI independently introduced contexts involving particular foods, animals, locations, and scenarios. This interaction illustrates how Teacher G retained the authority to accept or reject generated possibilities while delegating much of the generative authority for determining the specific personalization context to GenAI.

Teacher F gave the AI greater control selecting personalization references but asked it to leave the numbers in the problem and the mathematical structure of the problem alone. As he explained, “I thought about changing the numbers but, then I’m kind of like, I don’t want to compromise the math that they are doing.” Teacher F’s interactions complicate the characterization of AI-dominated personalization as entirely passive and time efficient. The teacher moves that were coded included a combination of both AI-determined specific personalization references (10.96% of teacher moves) and teacher-generated changes to personalization references (12.33% of teacher moves), as well as attention to contextual realism (16.44% of teacher moves), unrealistic numbers (16.44%), age of students (10.96%), and popularity among students (10.96%). Notably, Teacher F averaged 6.62 prompts per problem, the highest of the seven teachers, and spent an average of approximately 4.42 minutes per problem. Teacher F had slightly positive random intercepts for both Interestingness (0.104) and Closeness (0.034).

Teacher F’s interaction illustrates AI-Dominated Passive Prompt Personalization, in which greater generative decision-making authority was delegated to GenAI (Fig. 5). Rather than entering the interaction with specific personalization references already selected, Teacher F frequently provided broad criteria, such as asking GenAI to make the problem “relevant to Middle School students’ interests without specifying how, asking the GenAI to use a “popular city without giving guidance for how to select what is popular,” or asking the AI to situate the problem at “ a popular event that Middle schoolers are interested in,” leaving GenAI to determine the specific context and references that might satisfy those criteria. However, although he gave the GenAI a lot of brainstorming responsibility, Teacher F did not simply accept each generated response; he redirected the model, returned to earlier ideas, and selected among alternatives. And at times, when the AI model was not successful at giving him what he wanted through its own ideas, he became more directive (e.g., “Change this to the Winnebago County Fair”).

5.3 Human-AI Partnership

Teachers A and C demonstrated a human-AI partnership interaction style characterized by collaborative personalization, in which decision-making authority shifted between teacher and GenAI throughout the personalization process. Rather than maintaining control over nearly every personalization decision or relying primarily on GenAI to determine the context, these teachers selectively constrained features while allowing GenAI creative flexibility in other areas. Importantly, collaboration did not mean equal teacher and AI control at every stage. Instead, teachers determined when an AI output was useful and when teacher professional judgment was required to redirect, revise, or reject the GenAI output.

Teacher A demonstrated this movement across a broad range of coded interaction moves. The interactions included adding specific personalization references (17.54% of teacher moves), allowing GenAI to add specific personalization references (10.53% of teacher moves), independently swapping personalization references (12.28%), allowing GenAI to swap personalization references (10.53%), attending to contextual realism (17.54%), checking mathematical accuracy (10.53%), and prompting the AI to consider age level of students (14.04%). Rather than consistently keeping or letting go of control, Teacher A moved between these actions depending on the problem and Teacher A’s evaluation of the

generated output. Teacher A had the highest random intercept for Interestingness (0.173) and a strongly positive intercept for Closeness (0.167). Her case suggests that productive collaboration may involve maintaining teacher epistemic authority while selectively granting GenAI generative authority in areas where it can expand the range of personalization possibilities.

In Fig. 6, Teacher A demonstrated Human-AI Partnership through an iterative process in which she alternated between soliciting possibilities from GenAI and applying her own judgment to evaluate and refine those possibilities. This pattern was particularly evident when she transformed a problem about a baby giraffe's weight into a football context. Rather than specifying the complete personalization herself, Teacher A initially asked ChatGPT to "change this problem for me, to be about football instead," allowing the model to propose how the mathematical structure might be represented in the new context. She then evaluated successive responses on multiple dimensions, questioning the appropriateness of using a person's weight, whether the number of points scored was realistic, and whether the resulting solution would be an integer. When a revision produced 19 points per game, she reconsidered the change, noting that the value "may be a little low." She also asked ChatGPT to suggest an SEC team and subsequently narrowed that request to a team in a specific state before further refining how the scoring situation should be structured. Across these exchanges, Teacher A neither predetermined all the features of the personalized problem nor accepted GenAI-generated possibilities without evaluation.

Teacher C also demonstrated iterative refinement of word problems, but placed particularly strong emphasis on realism and mathematical appropriateness. Her coded moves included prompts related to contextual realism (19.42% of teacher moves), unrealistic numbers (18.45% of teacher moves), mathematical difficulty or clarity (16.50%), specific teacher-added personalization references (15.53%), attending to the age group of her students (10.68%), and modifying the readability of the problem (7.77%). Teacher C repeatedly evaluated GenAI outputs across mathematical, contextual, and student-centered dimensions. Teacher C's final interview described this iterative process. Teacher C stated that creating a successful personalization problem was "not just like a one-shot thing" and explained that "it takes a few times to refine it," particularly when checking whether prices, quantities, and context were realistic. At the same time, she recognized an important limitation of personalization: students "like what they like, and they don't like what they don't like." Students respond enthusiastically when contexts match their interests but were equally vocal when it did not. This observation is important because Teacher C's outcomes in terms of student ratings of personalized problems were less positive than those of A and B. Her random intercept for Interestingness was 0.021, while Closeness was − 0.059.

In Fig. 7, Teacher C also demonstrated Human-AI Partnership through iterative evaluation and refinement of GenAI-generated possibilities. Teacher C initially granted GenAI latitude to generate a context that was "trending with middle school students," but then repeatedly evaluated whether the emerging personalization was realistic and appropriate. As the context developed into designing a custom phone case, Teacher C refined how the quantities functioned within the scenario, used GenAI to investigate the typical size of an iPhone case, adjusted the numbers, and ultimately revised the problem for readability. This interaction illustrates how Teacher C moved between delegating generation and exercising

professional judgment, allowing GenAI to contribute possibilities while retaining responsibility for contextual, mathematical, and pedagogical appropriateness of the final problem.

Patterns across cases: From Control to Adaptive Epistemic Agency

The three interaction styles suggest that the allocation of control between teacher and GenAI alone did not explain students' perceptions of personalized mathematics problems. Productive teacher-GenAI collaboration may require teachers to maintain epistemic authority while adaptively allocating generative authority.

The strongest illustration is the contrast between Teacher E and G. Teacher E maintained substantial control, whereas Teacher G delegated substantially more personalization authority to GenAI. Despite occupying different end of the teacher-AI continuum, their student outcomes in terms of student ratings of personalized problems were remarkably similar: Teacher E had random intercepts of − 0.203 for Interestingness and − 0.191 for Closeness, while Teacher G had intercepts of − 0.200 and − 0.182, respectively. Thus, it may be that neither maximizing teacher authority nor maximizing AI authority is sufficient for producing personalized problems that students perceive as interesting and personally relevant.

6. Discussion

6.1 Teacher-GenAI Interaction Styles

Teacher-Dominated Rigid Personalization suggests that greater teacher control does not necessarily represent stronger epistemic agency. Wu et al. (2025) described adaptive epistemic agency as determining when to rely on, question, or redirect GenAI. In Teacher-Dominated Rigid Personalization, teachers retained substantial authority but provided GenAI limited opportunity to contribute to possibilities beyond teacher-determined changes. From an AI-TPACK perspective, this suggests that effective GenAI integration may require more than using AI to implement pedagogical decisions. Teachers must also recognize when its generative capabilities can productively extend their own knowledge and ideas (Ning et al., 2025).

AI-Dominated Passive Prompt Personalization illustrates the opposite tension: granting GenAI generative authority does not eliminate the need for teacher evaluation. Özer et al. (2026) argue that human epistemic authority depends on retaining responsibility for judging AI-generated outputs even when production is delegated to AI. This is particularly important for personalization because GenAI can generate context that appears relevant without necessarily reflecting the depth of specificity of students' actual interests (Bernacki & Walkington, 2026), which may help explain why generating an interest-related context alone did not necessarily result in students perceiving a problem as close to their interests. Therefore, delegating generation may expand personalization possibilities, but teachers remain responsible for evaluating their mathematical, contextual, and pedagogical appropriateness.

Collaborative Personalization most clearly reflects Wu et al.'s (2025) adaptive epistemic agency because teachers shift between directing, delegating, evaluating, and revising GenAI contributions. The finding extends this framework by distinguishing generative authority from epistemic authority: teachers can allow GenAI to generate contexts and possibilities while retaining responsibility for determining whether those possibilities are mathematically, pedagogically, and contextually appropriate. This also connects to TPACK and AI-TPACK, as teachers must integrate technological capabilities with mathematical, pedagogical, and student knowledge when evaluating GenAI-generated tasks (Mishra & Koehler, 2006; Ning et al., 2025). This finding also extends earlier work in which teachers used GenAI as a planning partner to generate multiple personalization possibilities while evaluating which ideas were appropriate for their students (Author, Date). The present study further conceptualizes this process as an adaptive allocation of authority, in which GenAI can contribute generative possibilities while teachers retain epistemic responsibility for evaluating and refining those possibilities.

6.2 Adaptive Allocation of Epistemic and Generative Authority

The findings suggest a useful distinction between generative authority and epistemic authority in teacher-GenAI collaboration. GenAI can be granted generative authority which is the latitude to propose contexts, references, quantities, and alternative representations without being granted epistemic authority over whether those possibilities should become part of the final instructional tasks. That responsibility remains with the teacher. This interpretation extends Wu et al.'s (2025) conceptualization of adaptive epistemic agency by emphasizing teachers' capacity to determine when to delegate, constrain, evaluate, or revise GenAI contributions rather than treating either teacher control or GenAI control as a primary way to personalize problems. Adaptive agency may involve not simply deciding whether to use GenAI, but continually determining which forms of authority can appropriately be delegated.

6.3 Limitations and Future Research

There are several limitations to this study to acknowledge. The number of teacher participants was 7; however, this study was in its first year, and 7 teachers were chosen to participate. The study includes pre- and post-interviews and problem-writing sessions 2 hours long, which provide detailed data logs of interactions with GenAI. Although the analysis used an established coding scheme and incorporated iterative within- and cross-case comparison of coded interactions and teacher reflections, future research could examine the robustness of the proposed interaction styles through independent coding. Finally, the quantitative data using teacher-level random intercepts for student ratings should be interpreted as exploratory only, given the small sample size.

6.4 Significance

Teachers don't simply apply a universal rule for using AI. They ask questions like: Is this context realistic? Is that number plausible? Will seventh graders recognize this person? Is this reference current? Did the context alter the mathematics? And importantly, should I change this myself or ask an LLM to generate

another possibility? Epistemic agency is visible in the conditional nature of those decisions. Our results suggest that genAI may help address longstanding practical challenges of personalization by making it more feasible for teachers to generate multiple versions of mathematical tasks, a process that has traditionally been time-consuming (Walkington & Bernacki, 2014). However, the ability to generate personalized problems at scale does not ensure meaningful personalization; teachers must still evaluate whether the generated contexts authentically and specifically connect mathematics to students' interests and experiences (Author, Date; Bernacki & Walkington, 2026).

This study contributes to emerging research on teacher-GenAI collaboration by showing that teachers do not simply choose whether to use or trust GenAI. Teachers need to make ongoing decisions about what to delegate, what to constrain, and what requires their professional judgment. The three interaction styles identified in this study-Teacher-Dominated Rigid Personalization, AI-Dominated Passive Prompt Personalization, and Human-AI Partnership Collaborative Personalization- demonstrate different ways teachers distribute responsibility during the design of personalized mathematics tasks. Importantly, findings suggest that productive teacher-GenAI interaction may depend less on how much control is held by the teacher of GenAI and more on how that authority is allocated throughout the design process.

A key contribution of this study is the distinction between *generative authority* and *epistemic authority*. This extends conceptualization of adaptive epistemic agency (Wu et al., 2025) by illustrating how teachers may enact that agency during mathematics instructional design. It also highlights that personalization with GenAI is not simply a matter of generating context connected to a student's interest; it requires teachers to evaluate the specificity of that connection. Personalization using GenAI in mathematics problem creation involves not just contextual relevance or knowing students' interests. It involves a teacher's epistemic stance, PCK, TCK, and, more importantly, using GenAI as a generative authority while teachers evaluate, iterate, and choose what is best for students. For mathematics teacher education, these findings suggest that preparing teachers to use GenAI should move beyond teaching prompt strategies. A teacher's capacity to generate, revise, and evaluate GenAI-generated outputs is of utmost importance as they maintain responsibility for the mathematical and pedagogical decision that ultimately shape a student's learning experience.

Declarations

AI Use Statement

The authors used a generative AI tool (ChatGPT) to assist with general editing of the manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contribution

Theodora Beauchamp: Methodology, Formal analysis, Writing – original draft, all figures and tables, Writing – review and editing. Candace Walkington: Conceptualization, Investigation, Methodology, Funding acquisition, Supervision, Writing – review and editing.Tiffini Pruitt-Britton: Writing – original draft; Writing – review and editing.

Acknowledgement

This work was supported by the National Science Foundation.

Data Availability

The qualitative data generated during this study (ChatGPT interaction logs and post-implementation reflections) are not publicly available due to participant confidentiality but are available from the corresponding author on reasonable request.

References

1. Aga, Z. G., Sawyer, A. G., & Wolfe, M. (2024). More confidence more critical: Investigating secondary mathematics preservice teachers' use of an Artificial Intelligence chatbot as a curriculum development tool. *American Journal of Creative Education*, *7*(2), 63–78. 10.55284/ajce.v7i2.1278
2. Author (Date).
3. Bernacki, M. L., & Walkington, C. (2026). An extensible model of personalized learning. In M. L. Bernacki, C. Walkington, A. Emery, & L. Zhang (Eds.), *Handbook of personalized learning* (pp. 46–60). Routledge. https://doi.org/10.4324/9781032719467-4
4. Biton, Y., & Segal, R. (2025). Learning to craft and critically evaluate prompts: the role of generative AI (ChatGPT) in enhancing pre-service mathematics teachers' TPACK and problem-posing skill. *International Journal of Education in Mathematics Science and Technology (IJEMST)*, *13*(1), 202–223. https://doi.org/10.46328/iemst.4654
5. Blundell, C. N., Mukherjee, M., Nykvist, S., et al. (2026). Teachers' TPACK Considerations Regarding Opportunities and Challenges of Using GenAI in K-12 Education. *Tech Know Learn*. https://doi.org/10.1007/s10758-026-10002-x
6. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
7. Cindilli, B., Sengul, B. Y., Kucuk, S., et al. (2026). Exploring gaps and determinants in teachers' artificial intelligence technological pedagogical content knowledge based on a cross-sectional study in Türkiye. *Discov Educ*, *5*, 402. https://doi.org/10.1007/s44217-026-01535-5

8. Creswell, J. W., & Poth, C. N. (2018). Qualitative Inquiry and Research Design: Choosing among Five Approaches. 4th Edition, SAGE Publications, Inc., Thousand Oaks.
9. Dogan, S. (2026). Designing effective AI professional development: A framework grounded in intelligent-TPACK. *Computers and Education: Open*, *10*, 100337. https://doi.org/10.1016/j.caeo.2026.100337
10. Drushlyak, M., Lukashova, T., Shamonia, V., & Semenikhina, O. (2025). ChatGPT-based simulation helps to develop the pre-service mathematics teachers' critical thinking. *International Journal of Instruction*, *18*(1), 153–172. https://doi.org/10.29333/iji.2025.1819a
11. Hussein, A., Daher, W., & Assali, A. (2026). Primary school teachers' I-TPACK competencies after gemini-based professional development. *Front Educ*, *11*, 1800871. 10.3389/feduc.2026.1800871
12. Karumbaiah, S., Yin, Y., & Brighouse, H. (2027). What Constitutes AI Harms and/or Unfairness? An Empirical Analysis of Teacher Deliberation with a Fairness Elicitation Scaffold. In E. G. Blanchard, G. Chen, M. Chi, & S. Isotani (Eds.), *Artificial Intelligence in Education. AIED 2026. Lecture Notes in Computer Science()* (Vol. 16586). Springer. https://doi.org/10.1007/978-3-032-29773-0_19
13. Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development*, *15*(3), 309–328. https://doi.org/10.1016/S0885-2014(00)00030-7
14. Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x
15. Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and Generative AI. *Journal of Digital Learning in Teacher Education*, *39*(4), 235–251. https://doi.org/10.1080/21532974.2023.2247480
16. Mosia, M., Nannim, F., & Egara, F. AI literacy-related domains and AI-TPACK readiness among preservice mathematics teachers: A factor-informed structural equation modelling study, Computers and Education: Artificial Intelligence, 11, 2026, 100650, ISSN 2666-920X, https://doi.org/10.1016/j.caeai.2026.100650
17. Ning, Y., Zhang, C., Xu, B., Zhou, Y., & Wijaya, T. T. (2024). Teachers' AI-TPACK: Exploring the Relationship between Knowledge Elements. *Sustainability*, *16*(3), 978. https://doi.org/10.3390/su16030978
18. Özer, M., Perc, M., & Özçelik, H. T. (2026). Human agency and epistemic authority under generative artificial intelligence. *Open Praxis*, *18*(3), 495–505.
19. Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models. Applications and Data Analysis Methods* (2nd ed.). Sage.
20. Seo, K., Yoo, M., Dodson, S., & Jin, S. H. (2024). Augmented teachers: K–12 teachers' needs for artificial intelligence's complementary role in personalized learning. *Journal of Research on Technology in Education*, 1–18. https://doi.org/10.1080/15391523.2024.2330525
21. Uğraş, H., Uğraş, M., Papadakis, S., & Kalogiannakis, M. (2024). ChatGPT-Supported Education in Primary Schools: The Potential of ChatGPT for Sustainable Practices. *Sustainability*, *16*(22), 9855.

https://doi.org/10.3390/su16229855

22. Author (Date).
23. Author (Date).
24. Wu, H., Ning, Y., Chang, H., et al. (2026). AI-TPACK Competencies in Mathematical Problem-Posing Teaching: Measurement Instrument Development and Latent Profile Analysis of Pre-Service Teachers. *Int J of Sci and Math Educ*, *24*, 41. https://doi.org/10.1007/s10763-026-10666-y
25. Wu, J. Y., Lee, Y. H., Chai, C. S., & Tsai, C. C. (2025). Strengthening Human Epistemic Agency in the Symbiotic Learning Partnership With Generative Artificial Intelligence. *Educational Researcher*, *0*(0). https://doi.org/10.3102/0013189X251333628
26. Xie, M., Luo, L., & Assessing AI-TPACK readiness in mathematics teacher education: The role of self-efficacy and teaching beliefs, Computers and Education Open, 10, 2026,100375, ISSN 2666–5573, https://doi.org/10.1016/j.caeo.2026.100375

Figures

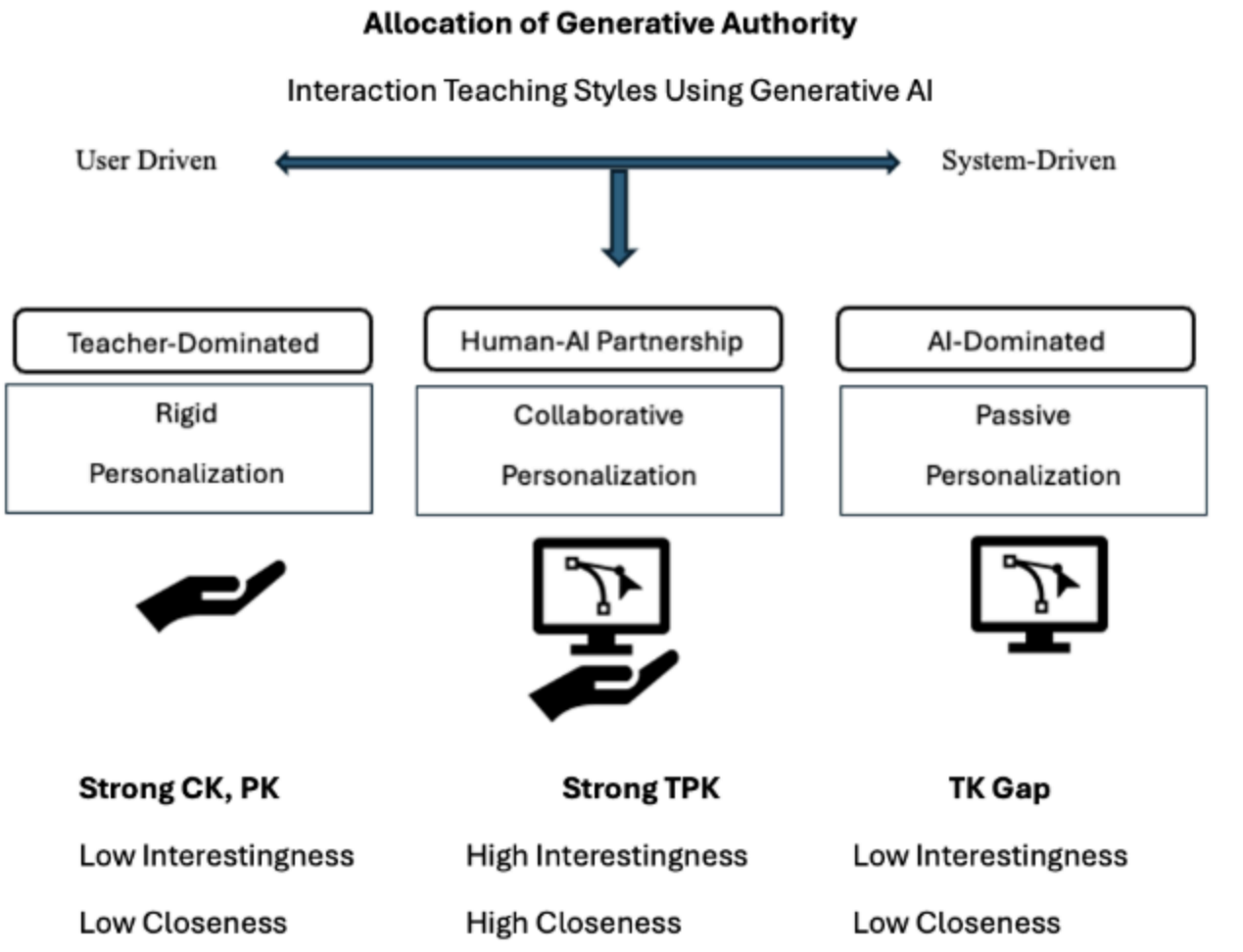


Figure 1

Interaction Teaching Styles Using GenAI

Original Problem

Andre wants to buy a backpack. The normal price of the backpack is $40. He notices that a store that sells the backpack is having a 30% off sale. What is the sale of the backpack?		
Teacher D prompt	**Actual Code(s)**	**Analytic Description**
"Can you use this math problem to make a more relevant problem for my 7th grade students who need to buy something for soccer?"	Initial; Add general PC topic; age	Establishes the personalization direction. Specifies soccer and Grade 7 students as parameters for GenAI's initial output.
"Can we change the name Xander to Jayden?"	Names	Specifies student name. Replaces GenAI's generated name with a teacher-selected name.
"Can you add the name brand Nike to the soccer cleats?"	Add specific PC reference	Adds a teacher-selected specific reference.
"Can you change the price of the cleats to $70."	Not realistic-number	Specifies numerical realism. Directly determines the price used in the personalized context.
"Can you combine the first two sentence?"	Other	Directs wording and structure. Specified how GenAI should reorganize the generated problem.
""Can you add that he is buying them from Academy?"	Add specific PC reference	Adds a teacher selected store reference. Further specified the personalization context.
"Can you remove the word normally?"	Other	Refines wording. Directly identifies language to remove from the problem.
"Move the name of the store to the second sentence."	Other	Refines sentence organization. Directs the placement of a specific contextual detail.

Final output

Jayden wants to buy a new pair of Nike soccer cleats, which cost $70. He notices that Academy is having a 30% off sale on the cleats. What is the sale price of the Nike soccer cleats?

Figure 2

Teacher D's Rigid Personalization Process. PC stands for "popular culture."

Original Problem

The price of gold is often reported per ounce. At the end of 2005, this price was $513. At the end of 2015, it was $1060. By what percentage did the price per ounce of gold increase?		
Teacher E prompt	**Actual Code(s)**	**Analytic Description**
"Can you rewrite this relating to relating to number of songs on Spotify?"	Initial; Add specific PC reference	Establishes the personalization context. Teacher specifies Spotify and song quantity as the direction for personalization.
"Nice try. Can you take out million and make it specific to rap music?"	Add specific PC reference; Other	Further constrains the personalization. Teacher rejects part of the generated version and specifies rap music as the desired reference.
"Nice try. Can you change the date to a current year?"	Other	Directs revision toward numbers. Teacher identified a temporal feature that should be changed.
"Nice try, can you make the date 2023 to 2024?	Not realistic-context	Specifies the exact timeframe. Teacher replaces the broader request for a current year with precise dates.
"Nice try, can you use a sub genre of Rap?"	Not realistic-number	Narrows the musical context. Teacher directs GenAI toward a more specific personalization reference.
"Nice try, can you use a sub genre of Pop?	Not realistic-context	Redirects the previously selected personalization. Teacher replaces the rap direction with a pop subgenre.
"Nice try, can you rewrite using a one-month time frame?"	Not realistic-number	Constrains the structure. Teacher specifies a shorter timeframe for the generated situation.
"Nice try, can we use Beyonce over a 10-year time span?"	Add specific PC reference; Not realistic-context	Specifies both reference and timeframe. Teacher selects Beyoncé and replaces the one-month timeframe with a 10-year period.

Final output

At the end of 2013, there were 513 Beyoncé songs available on Spotify. By the end of 2023, this number had increased to 1,060 Beyoncé songs. By what percentage did the number of Beyoncé songs on Spotify increase?

Figure 3

Teacher E's Rigid Personalization process. PC stands for "popular culture."

Original Problem

"A 50-centimeter piece of wire is bent into a circle. What is the area of this circle?"		
Teacher G prompt	**Actual Code(s)**	**Analytic Description**
"Can you make this problem about food?"	Initial; Add general PC topic	Delegates broad context generation. Teacher G establishes "food" as the general personalization topic while leaving GenAI to determine the specific food reference.
"Can you try a different food?"	Swap specific references (LLM); Not realistic-context	Rejects generated reference while delegating the alternative. Teacher G evaluates the generated context as unsuitable but does not specify a replacement, leaving GenAI to generate another possibility.
"How about a different one?"	Swap specific reference (LLM)	Delegates alternative generation. Teacher rejects the current reference without identifying a desired alternative, returning generation of the specific personalization reference to GenAI.

Final output

A 50-centimeter strip of fruit roll-up is shaped into a circle. What is the area of the circle made from the fruit roll-up?

Figure 4

Teacher G AI-Dominated. PC stands for "popular culture."

A phone keeps track of the number of steps taken and the distance traveled. Based on the information in the table, is there a proportional relationship between the two quantities? number of steps distance in kilometers 950 1 2,852 3 4,845 5.1 Explain your reasoning.		
Teacher F prompt	**Actual Code(s)**	**Analytic Description**
"Rewrite this problem to make it relevant to Middle School student's interests."	Initial; age	Delegates initial personalization generation. Establishes relevance to middle-school students as the criterion but leaves GenAI to determine the specific personalization context.
"Use this same problem but don't use steps taken and distance traveled."	Add general PC topic	Redirects the generated context broadly. Requests a different context without identifying what the replacement should be.
"Go back to steps taken."	Not realistic-context; regeneration request	Rejects the alternative and return to the original mathematical context.
"Give the fitness app a name and use apple watch instead of smartphone."	Add specific PC reference (LLM); Swap specific PC reference	Combines delegation and teacher specification. Ask GenAI to generate the app name while directly specifying Apple Watch.
"Give me the original problem."	Add specific PC reference (LLM)	Returns to the original task as a new starting point.
"Give the phone a name."	Add specific PC reference (LLM)	Delegates specific-reference generation. Leaves selection of the phone reference to GenAI.
"Use a different phone name."	Add specific PC reference (LLM); Not realistic-context	Rejects the generated reference while delegating its replacement.
"Give this a location."	Add specific PC reference (LLM)	Delegates location generation. Requests contextual specificity without supplying the location.
"Now use it in a popular city."	Swap specific reference (LLM); location; popularity	Adds evaluative criteria while leaving the specific choice to GenAI.
"Take location out."	Swap specific PC reference	Rejects the location-based personalization direction
"At a popular event that Middle schoolers are interested in."	Add general PD topic; age; popularity	Establishes relevance criteria while delegating event selection
"Try again but start with the original problem I gave you. I need it to be something relevant that students will want to engage with."	Not realistic-context	Rejects the current iteration and reasserts student relevance as the goal.
"Change this to the Winnebago county fair."	Add specific PC reference; location	Teacher selects the specific context.
"Change phone to a phone brand."	Swap specific reference (LLM)	Delegates generation of a more specific phone reference.
"Give me the latest Galaxy model instead."	Swap specific PC reference	Refines the reference by specifying a more current model.

Final output

At the Winnebago County Fair, a Samsung Galaxy S24 tracks the number of steps taken and the distance traveled as students explore food stands, rides, and exhibits. Based on the information in the table, is there a proportional relationship between the two quantities? number of steps distance in kilometers 950 1 2,852 3 4,845 5.1 Explain your reasoning.

Figure 5

Teacher F's AI-Dominated Passive Prompting. PC stands for "popular culture."

Original Problem

"This is a problem I have for math. The baby giraffe weighed 132 pounds at birth. He gained weight at a steady rate for the first 7 months until his weight reached 538 pounds. How much did he gain each month? Could you change this problem for me to be about football instead."		
Teacher A prompt	**Actual Code(s)**	**Analytic Description**
"This is a problem I have for math. The baby giraffe weighed 132 pounds at birth. He gained weight at a steady rate for the first 7 months until his weight reached 538 pounds. How much did he gain each month? Could you change this problem for me to be about football instead."	Initial; Add general PC topic	Opens personalization by introducing a broad football context
"Okay. Could you do something maybe without using the weight of a person?"	Not realistic-context; Other	Rejects an inappropriate context
"Is this number a little over what a team would actually score in a season?"	Not realistic-number	Evaluates numerical realism
"yes, please make it more realistic."	Not realistic-context	Requests revision for contextual realism
"That still seems kind of high after 7 games. If we are making about college teams, could you make it a little lower and realistic."	Not realistic-context	Continues evaluating realism after GenAI revision
"Okay. Can we make sure the solution is an integer?"	Math Difficult or Clarity	Evaluates mathematical structure of the personalized problem
"Thanks. Can we maybe make it go back to the last numbers you used because 19 points per game may be a little low."	Not realistic-context	Reconsiders the revision based on plausibility
"Okay. Can you name the team for an SEC team?"	Add specific PC references (LLM); location	Narrows GenAI-generated reference using location
"Can we be more specific as in they started with a number of points as the beginning of the season and to ask the number of points per game the second half of the season?"	Math Difficulty or Clarity	Refines how the personalized context represents the mathematics

Final output

The (specific university) scored 133 points in the first half of the season. Over the next 6 games, they steadily increased their total points until they had 265 points by the end of the season. On average, how many points did they score per game in the second half of the season?

Figure 6

Teacher A Human-AI Partnership

Original Problem

Elena makes her favorite shade of purple paint by mixing 3 cups of blue paint, 1 1/2 cups of red paint, and 1/2 of a cup of white paint. Elena has 2/3 of a cup of white paint. Assuming she has enough red paint and blue paint, how much purple paint can Elena make? How much blue paint will Elena need to use with the 2/3 of a cup of white paint? How much red paint will Elena need to use with the 2/3 of a cup of white paint?		
Teacher C prompt	**Actual Code(s)**	**Analytic Description**
"Can we rewrite this problem to talk about something that is trending with middle school students."	Initial; age; popularity	Delegates initial context generation. Establishes student relevance and popularity as criteria while leaving GenAI to generate the specific context.
"Can we try a different story with something trending online?"	Add general PC topic; popularity	Rejects and asks for new option. Delegates GenAI to generate another possibility.
"Can we rewrite this to make a story realistic to what a middle school student would do?"	Not realistic-context	Evaluate contextual authenticity. Uses knowledge of middle-school students to constrain the developing personalization.
"Can we change sheets of vinyl to different categories of pictures to add to their phone case?"	Not realistic-context	Redirects the GenAI-generated context. Teacher C contributes a specific modification to make the emerging phone-case scenario more plausible.
"Can we change the number of pictures to the space the pictures take up on the phone case?"	Not realistic-quantity	Evaluates quantity/contextual fit. Revises how the quantities function within the personalized situation.
"Can we change the color theme to categories like family, friends, and pets?"	Not realistic-context	Refines contextual meaning. Teacher C replaces an abstract color-theme structure with personally meaningful picture categories.
"How many square inches is a typical iPhone case?"	Check accuracy	Uses GenAI to investigate realism. Seeks information to evaluate whether the generated quantities are plausible in the new context.
"Yes, can you adjust it to an iPhone 13/14/15 size?"	Not realistic-number	Applies the realism check to revision. Uses the information generated through the interaction to constrain the numerical context.
"Can we rewrite this but double all the numbers."	Math difficulty or Clarity	Reasserts mathematical requirements. Modifies the quantities while preserving the personalized context.
"Can we rewrite this with shorter sentences."	Readability	Evaluates pedagogical accessibility. Refines the final problem for student readability.

Final output

A dance team is practicing a routine with different styles of music. They planned to spend 3 minutes on hip-hop, 1 1/2 minutes on jazz, and 1/2 minute on contemporary. They actually spent 2/3 of a minute to practice the contemporary section. If they practice the other sections at the same rate, how much total time will they need for the routine? How much time will they spend on hip-hop if they use 2/3 of a minute for contemporary? How much time will they spend on jazz?

Figure 7

Teacher C Human-AI Partnership

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix.docx