Artificial Intelligence and Cognitive Demand During a Domain and Range Mathematics Task

Victoria Delaney
San Diego State University

Lily Sawi
San Diego State University

Sunday Stein
San Diego State University

Ryan Aniceto
San Diego State University

This study examines students' cognition during high demand mathematics tasks when an AI tool is available. Using video recordings, we analyze 20 undergraduates' cognitive demand on a task involving a nonroutine roller coaster application of domain and range. Our findings show wide variability in how students approach the task, how and how often they elected to use ChatGPT (half chose to use it, half did not), and the levels of cognitive demand that students exhibit in their solutions and thinking. Students that solve the task with high demand tend to grapple with its nonroutine aspects and challenge AI output, while students that solve with low demand give decontextualized prompts, struggle to use AI output, and/or trust AI despite hallucinations. We discuss implications that undergraduate mathematics researchers and educators might consider to preserve opportunities for students to experience high demand problem solving.

Keywords: Domain and range, cognitive demand, AI in education

Artificial intelligence (AI) tools are becoming commonplace in higher education. Although the notion of using computers to support, transform, or even replace mathematical thinking is not new (Kaput, 1992; Pea, 1985), AI tools present a new frontier for teachers and students as they can mimic human reasoning at most stages of problem solving. Prior and recent works that have studied AI in mathematics education have reached varying conclusions. For instance, AI can serve as a helpful scaffold when used deliberately (Yoon et al., 2024) and in ways that preserve human demand for thinking. On the other hand, AI tools can stunt learning opportunities when students use them in ways that disable their ability to master mathematics on their own (Bastani et al., 2024). Regardless, findings from these works suggest that ignoring the impact of AI tools on students' thinking and learning opportunities will pose serious detriments to researchers and mathematics practitioners. It is our goal to draw attention to this phenomenon by studying the nature of students' mathematical thinking when AI tools are made available to them.

Examining how students think with AI tools matters because AI prompting is a cognitive activity. When querying an AI chatbot, a student must *envision* what it is they need to know and write a query that will likely generate the information that they need to reach resolution (Subramonyam et al., 2024). At present, we lack research that details the extent to which students (a) engage with mathematical envisioning as they create AI prompts, and (b) how the presence of AI impacts their own cognition as they reason mathematically on demanding tasks. This study addresses both of these gaps. Using video analysis and attending carefully to students' thinking during AI interactions (e.g., Stein & Delaney, 2025), we observed 20 undergraduate students solve high demand tasks when given access to ChatGPT. We ask, *what is the nature of the cognitive demand when students solve a nonroutine mathematics task with AI?*

Theoretical Framework

We examine the changes in cognitive demand during a mathematics task and the interactions between students and AI throughout its enactment. Stein and colleagues refer to a *mathematics* *task* as “activity, the purpose of which is to focus students' attention on a particular mathematical idea” (*note:* we omit the “classroom” aspect of their definition as our study was conducted with individual students) (1996, pg. 460). Historically, mathematics tasks have been considered in the context of what students are expected to produce with a given set of resources (Doyle, 1983). AI tools may potentially change this conception, since they function as general-purpose cognitive resources that remove the need for focused student attention, depending on how they are used.

High demand tasks have functioned as a primary source from which students can learn (Shavelson et al., 1986). These tasks demand cognitive processes, such as identifying key information and activating prior knowledge that is useful for constructing solutions (Doyle, 1983). *High cognitive demand tasks* require students to make conceptual connections, generate nonalgorithmic solutions, and consider multiple solution strategies (Boston, 2012; Stein & Smith, 1998). Again historically, technologies have influenced mathematical activity and cognition, including the cognitive demand of tasks (Hollebrands, 2003; Sherman, 2014). However, technologies’ potential to mediate student learning has been contingent on students’ knowledge about how to use them productively toward the task’s goals (Delaney & Kinsey, 2021). As students can use AI tools without sophisticated knowledge of how they work, it is not yet known if, for whom, and to what extent AI tools will influence students’ cognitive demand as they solve rigorous math tasks.

Methods

Four researchers with expertise in undergraduate lower-division mathematics education and AI tool integration designed and collected data for this project in spring 2025. Our goal was to investigate the nature of students’ mathematical thinking and reasoning when ChatGPT-4o was made available to solve a mixture of closed-form and non-routine questions. The questions were given to students in three separate 20-minute tasks. In this paper, we analyze 20 students’ responses to one component of the algebra task (see Figure 1), and the ways in which students’ uses of ChatGPT impacted the cognitive demand of the task as they solved it.

We recruited students via physical campus advertisements. To participate, students needed to have taken a mathematics course in the past 12 months to reduce the likelihood that they would be completely unfamiliar with the tasks’ mathematics concepts. We also collected information about students’ academic year, major, and undergraduate math courses. We recruited from a variety of majors, and 9 of the 20 participants were not explicitly STEM students, although some listed 4-5 math intensive courses on their screener survey. They studied biology, geography, liberal studies (2), marketing, business management, economics, food/nutrition, and psychology.

During a 10-minute pre-task interview, one researcher asked participants additional questions about their typical AI use on school assignments, beliefs about AI, and the extent to which their instructors demonstrated or encouraged AI use in their courses. A detailed account of students’ AI beliefs and practices can be found in (Delaney et al., 2025). Following the pre-interview, each participant was instructed to “think aloud” as they solved the math tasks on an iPad screen that recorded their responses temporally. A lead interviewer reminded students to think aloud, asked open-ended questions (e.g., “what are you noticing right now?”) during problem solving, and assured participants that we were primarily interested in studying their thinking and reasoning. A second researcher managed video and audio recordings, kept time during the sessions, and took field notes that captured instances where the student appeared to be grappling with the mathematics. We concluded with a post-task 10-minute debrief.

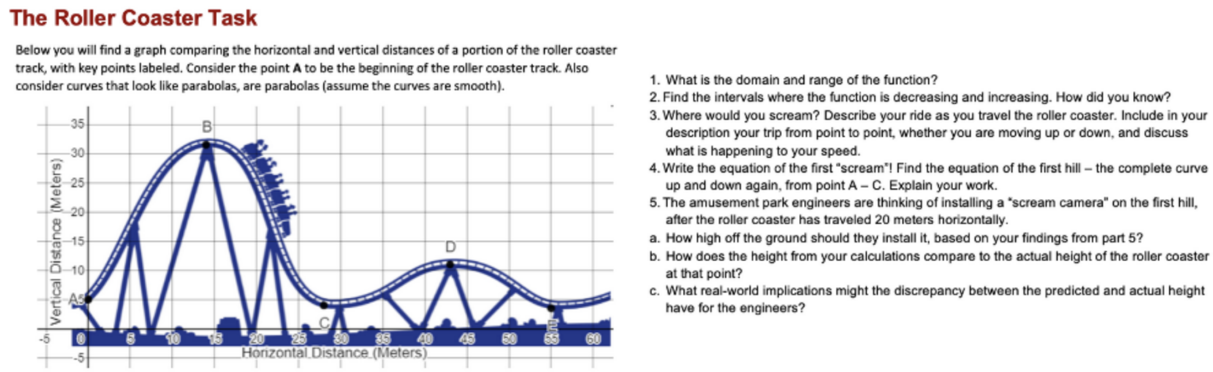

Figure 1. The roller coaster task, adapted from the Virginia Department of Education.

Data Analysis

We focus on the domain and range subcomponent of the Roller Coaster task because it is a routine question (*"What is the domain and range of the function?"*) with a nonroutine solution. We conjectured that if students maintained the cognitive demand of the task, they would attend to the ambiguous right-side of the graph (a high demand feature), which continues despite a fifth point labeled "E" at around (4, 55). We further hypothesized that students' *in situ* experiences with roller coasters would enable them to reason about its domain and range holistically, rather than anchoring on canonical domain/range representations and processes from math courses.

The first and second authors analyzed the Roller Coaster task video recordings (Derry et al., 2010) of each student using the Mathematics Task Framework (Boston, 2012; Stein & Smith, 1998). We contextualized the Mathematics Task Rubric to the Roller Coaster task, first solving it ourselves and mapping students' solutions to definitions from the original framework (see Table 1). We individually rated students' cognitive demand as they gave their solutions, noting that some students experienced higher levels of cognitive demand as they were thinking about domain and range compared to their demand as they wrote down their domain/range solution. After giving our initial ratings, we collectively compared ratings, discussed our interpretations with respect to the framework, and resolved disagreements. For example, we determined criteria for "complex, nonalgorithmic thinking" and reviewed the ratings we gave iteratively to ensure that we were interpreting them consistently. Once initial agreement was reached and ratings stabilized, the first author reviewed all ratings one final time to check for internal consistency.

Table 1. Cognitive demand rating of the roller coaster task, adapted from the Mathematics Task Rubric

Definition	Domain Definition and Example	Range Definition and Example
0 - Not applicable *Guessing, non-math thinking*	Guessing without procedures or mentioning previously learned domain facts, rules, definitions.	Guessing without procedures or mentioning previously learned range facts, rules, definitions.
1 - Memorization *Reproducing previously learned facts, rules, or definitions.*	Reproducing previously learned values for the domain. ***Example:*** "I would say negative to positive infinity…it was 100% on the AP Calc test."	Reproducing previously learned values for the range. ***Example:*** "I can remember from high school…range just keeps going, unless it ends at zero."

2 - Procedures without Connections *Focused on correct answers without connecting to underlying meaning.*	Finds values for the domain without connecting to underlying meaning or ambiguity of where the track ends. ***Example:*** "So the domain is like the x, so it goes from 0 to 60 meters."	Finds values for the range without connecting to underlying meaning or grappling with multiple local nonzero minima. ***Example:*** "The range is the y, so it goes from zero to…34? 32.5? This is the maximum, it's not quite 35."
3 - Procedures with Connections *Procedures for the purpose of developing deeper levels of understanding of mathematical concepts.*	Reasons about domain by connecting to ambiguity about where the track ends. ***Example:*** "...Assume it starts at zero, there's a closed endpoint …and it goes up until 60? 55? I'm going to say infinity, because it goes on with nothing else."	Reasons about both range values (starting and ending) in the context of the graph's min. and max. ***Example:*** "It looks like the lowest point is at 5 meters, and the highest is at the peak, point B, which looks like 32 meters. So yeah, 5 to 32."
4 - Doing Mathematics *Complex, nonalgorithmic thinking, use relevant knowledge and experiences.*	Reasons about the domain in the situated context of a roller coaster. ***Example:*** "It says consider point A to be the start. But the track has to connect with this, eventually, come around to point A."	Reasons about the range in the situated context of a roller coaster. ***Example:*** "B is the highest peak on the graph. There could be a higher peak beyond there, but at least on this graph, the peak is 33."

Findings

Students' strategies, ChatGPT uses, and cognitive demand varied drastically. Most students spent 0:30-4:00 minutes solving for domain and range, although some grappled with its meaning for 4-10 minutes and one aerospace engineering student spent all 20 minutes on it. Figure 2 shows students' cognitive demand, the time they spent, and whether or not they used AI. Table 2 shows students' written responses to the task, if they used AI, and how many prompts they used. Completely coincidentally, 10 students used ChatGPT and 10 did not.

*Table 2. Students' written solutions to the domain and range portions of the roller coaster task. Students who chose to use ChatGPT are listed in the left two columns with number of prompts given in parentheses. *Indicates STEM.*

Student (no AI)	Written Solution	Student (used AI)	Written Solution
Student_1	D: $(-\infty, \infty)$, R: [5, 30]	Student_4 (1)	D: n/a, R: $3 \leq y \leq 32$
Student_2	D: (0, 55), R: (5, 32)	*Student_5 (1)	D: [0, 55], R: [4, 32]
Student_3	D: $(0, \infty)$, R: (5, 32)	*Student_6 (5)	D: (0, 55), R: (5, 30)
Student_9	D: (0, 60), R: (0, 35)	Student_7 (3)	D: [5, 10], R: $[5, \infty)$
*Student_10	D: (0, 63), R: (5, 32)	Student_8 (3)	D: $(0, \infty)$, R: (5, 32)
*Student_11	D: $[0, \infty)$, R: [4, 32]	Student_13 (2)	D: $(-\infty, \infty)$, R: (4, 32)

*Student_12	D: {x | x ≥ 0), R:{y | y ≥ 0)	*Student_14 (1)	D: [0, ∞), R: [3, 33]
*Student_15	D: [0, ∞), R: [3, 32]	*Student_16 (1)	D: (0, ∞), R: (5, 32)
*Student_18	D: [0, 60], R: [0, 32]	Student_17 (8)	D: [5, 55], R: [-5, 35]
*Student_20	D: [0, 62.5], R: [4, 32]	*Student_19 (1)	D: [0, ∞), R: [0, 32]

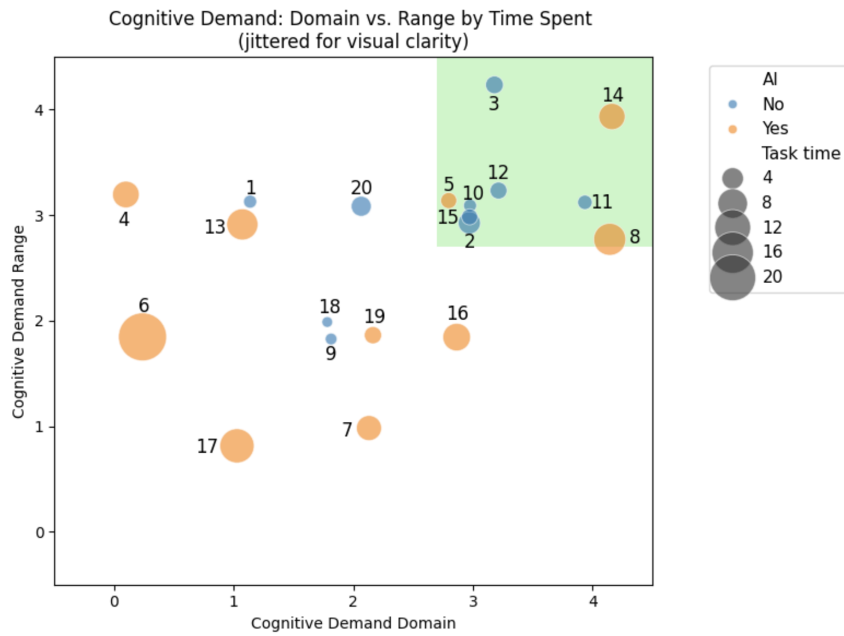

Figure 2. Cognitive demand (domain and range), time spent solving, and AI use of 20 student participants. The numbers in the figure indicate the student ID number (as listed in Table 2).

High Cognitive Demand

9 of 20 students maintained high cognitive demand for both domain and range (a 3 or 4 on the Mathematics Task Rubric, Stein & Smith, 1998), indicated by the green shaded region in Figure 2. A majority of these students (6 of 9) did not use ChatGPT to assist them with domain and range. Rather, they tended to remember what domain and range were and had ideas about how to apply them in the novel roller coaster setting. If they did use ChatGPT, it was typically a single prompt to remind themselves which axis domain or range corresponded to (e.g., Student_14 prompted, *"Does range measure the y axis?"*). They noticed at least one discrepancy between the plotted points (A-E) on the roller coaster track, the track visually continuing beyond the figure boundaries, and they generated questions pertaining to the domain's ambiguity about where the track ended (and *if* it ended). To illustrate, Student_8, a business management major who first used ChatGPT to recall domain and range concepts, said:

> "But you can see it continues this way. But [*in the problem statement*] it says consider point A to be the beginning of the roller coaster track. But this (*gestures to the track*) has to connect to the beginning eventually…come around to point A. So this makes me think that the domain is all real numbers. But at the same time, a roller coaster is not infinite."

Here, Student_8 explained her struggle to find a reasonable endpoint for the domain of the roller coaster. She knew it could not be positive infinity, because the roller coaster "has to connect eventually", yet she was uncertain what number would make sense, apart from numbers visualized on the graph itself, which could not be the domain's endpoint because the roller coaster continued past them. The nature of her struggle showed high demand thinking (level 4, doing mathematics) because she grapples with non-algorithmic solution strategies, such as attempting to visualize the rest of the non-infinite roller coaster.

Student_8 and others struggled to generate a reasonable endpoint for the roller coaster's domain. As shown in Table 2, students created one of three responses: 55 (the x-coordinate of point E), 60-63 ("some point beyond 55"), or positive infinity, even though many students who wrote ∞ stated aloud that the roller coaster could not go to infinity. Student_16, a mechanical engineer, explained this discrepancy in terms of how she would approach it in her major:

> "I know that the roller coaster does not go on forever, but I have no information to suggest where it will end. So I mean, in the context of engineering, if this goes on for…65, 60 meters of it, and it's a 1000-meter roller coaster, you theoretically would consider that infinity, because it's so much higher. But I just don't have that information."

Low Cognitive Demand

11 of 20 students solved the task with low cognitive demand for domain, range, or both (typically it was domain - see Figure 2). Of these 11, 5 were STEM majors and 6 were not, suggesting that students with more math exposure through coursework did not appear to solve the task while maintaining higher cognitive demand than their non-STEM peers.

Students who solved with low cognitive demand tended to not grapple with the situated ambiguity presented by the roller coaster. For instance, Student_18 (an electrical engineering major) spent only 39 seconds on the task, reading the graph and immediately declaring, "So the domain is x, and it goes from 0-60 meters. The range is y, and it goes from 34? 32.5 meters. Because this is the maximum." He did not explain where 60 meters came from, but rather scanned the graph and selected domain values (level 2, procedures without connections). Student_13 (a biology major), on the other hand, spent almost nine minutes puzzling about the domain, before concluding, "Am I allowed to guess on this? My guess would be from negative infinity to infinity…it was 100% on the AP Calc test…if you were given a parabola and had to determine all these things about it." (level 1, memorization).

7 out of 11 students used ChatGPT to solve the task, but only three used it in ways that led them to generate a high demand answer for either domain or range (Student_4 and 13 for range, and Student_16 for domain). The others, and in particular Student_6, Student_17, and Student_7, struggled to obtain helpful information about the task via time intensive back-and-forth interactions with ChatGPT. Their struggle to use ChatGPT productively originated from one of three potential causes. First, they did not know how to make use of the information ChatGPT gave them, because it gave generic facts and examples about parabolas' domains and ranges in response to students' generic questions (Student_13's prompt: "determine domain and range on a parabola."). Student_13 struggled to relate ChatGPT's hints about utilizing a function in vertex form, to the roller coaster, which did not give a function equation:

> "It gave me the equation of a parabola, then it put it into vertex form…and then h and k is the vertex…And then it tells me the domain, based on, plugging in x and getting any real number. I mean, it's kind of clarifying, but I don't have the, like, function for it."

Other students, such as Student_7 (a food and nutrition major), interpreted ChatGPT's hints incorrectly. Student_7, reading from ChatGPT, inferred that domain is, "where it starts and where it finishes," but applied this definition in such a way that she obtained two answers, "and I want to say that the domain would be x=5 and x=10." Her response reflects an understanding that a domain is discrete values, which she explained came from ChatGPT's illustrative example.

Finally, although hallucinations did not appear to occur frequently, Student_17 struggled when ChatGPT gave her negative range values after she prompted it with a screenshot of the roller coaster from the task. Student_17, a liberal studies major, realized that the graph visualization did not contain negative values, but did not know how to prompt ChatGPT in such a way that it would produce positive range values:

> (*reading AI output*) "y is the line that goes up and down, from the lowest point to the highest, drops as low as -5 meters and climbs to 35 meters. (*stops reading*). No it doesn't. Because, right here (*points to B, vertex*), looks like it's maybe 32. So, wrong, incorrect."

However, Student_17 was unsure how to find the domain and range absent a visually logical response from ChatGPT. After about 9 minutes and eight prompts, she resigned, "Okay, I'll just say what it tells me" (see Table 2), even though she suspected that ChatGPT's response was incorrect ("I don't know what -5 has to do with it") and had her own emerging ideas about how to find the range without AI (" *positive*) five is on this line, at the same place as zero").

Discussion

Consistent with prior work that has studied technology as a mediator of cognition (e.g., Hollebrands, 2003; Sherman, 2014), our findings suggest that AI tools do not, in many cases, deterministically lower or raise cognitive demand. Rather, how students solve the task depends on how they interpret it, construct prompts that bridge their understanding with what they need to learn, and evaluate AI output in response to the task's goals. Importantly, we observed that evaluating AI involved students' willingness and ability to challenge ChatGPT as a technological form of authority. Student_8 did so successfully, using it to remember domain and range then moving toward the demanding, ambiguous aspects of the task herself, while Student_17 did not, writing down ChatGPT's incorrect solution as her own even though she did not agree with it.

What students wrote down as their domain/range solution did not unilaterally represent their thinking during the task. Many students, for instance, did not know how to turn "this roller coaster cannot go on forever" into a domain statement even though they exhibited high demand thinking. Furthermore, ChatGPT did not tend to provide clarification because it gave canonical parabola domain examples that did not apply to the roller coaster. Rather than follow up with prompting that alluded to a need to *envision* a roller coaster as a function (Subramonyam et al., 2024), students with low demand thinking translated ChatGPT's generic output into ideas they believed were usable, often subverting qualities that made the task high demand in the first place.

These findings hold implications for the undergraduate mathematics education community. First, we hope that our findings illustrate that some students need to learn how to use AI tools productively to solve problems by becoming more attuned to their own sense-making and discernment. Second, educators must help students learn how to navigate open-ended, ambiguous problems, particularly as AI tools become more accurate in solving routine math problems (Pardos & Bhandari, 2024). Finally, as our study was conducted with individuals outside of any course they were enrolled in, future work should explore the nature of students' cognition and AI use within courses, where their AI use habits may have a stronger impact on their grades and eventual career trajectories. Future work might also consider how to design student-AI interactions, particularly with frontier AI models that students use on a regular basis, so that student thinking is amplified rather than automated.

References

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, O., & Mariman, R. (2024). Generative AI can harm learning. *Available at SSRN*, *4895486*.

Boston, M. D. (2012). Assessing instructional quality in mathematics. *The Elementary School Journal, 113*(1), 76–104. https://doi.org/10.1086/666387

Delaney, V. & Kinsey, G. (2021). Sustaining Cognitive Demand with Desmos Technology. In *Proceedings of the 43rd Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, pp. 1764-1768.*

Delaney, V., Stein, S., Sawi, L. & Hernandez Holliday, K. (2025, October). Undergraduate students' appraisals and rationales of AI fairness in higher education. In J. Wilson, C. Ormerod, & M. Beiting Parrish (Eds.), *Proceedings of the Artificial Intelligence in Measurement and Education Conference (AIME-Con): Full Papers* (pp. 329-336). National Council on Measurement in Education. https://aclanthology.org/2025.aimecon-main.35

Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., Hall, R., Koschmann, T., Lemke, J. L., Sherin, M. G., & Sherin, B. L. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *Journal of the Learning Sciences, 19*(1), 3–53. https://doi.org/10.1080/10508400903452884

Doyle, W. (1983). Academic work. *Review of educational research*, *53*(2), 159-199.

Hollebrands, K. F. (2003). High school students' understandings of geometric transformations in the context of a technological environment. *The Journal of Mathematical Behavior*, *22*(1), 55-72. https://doi.org/10.1016/S0732-3123(03)00004-X

Kaput, J. J. (1992). Technology and mathematics education. In D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning , (pp. 515 –556). New York: MacMillan Publishing Company.

Pardos, Z. A., & Bhandari, S. (2024). ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills. *Plos one*, *19*(5), e0304013. https://doi.org/10.1371/journal.pone.0304013

Pea, R. D. (1985). Beyond amplification: Using the computer to reorganize mental functioning. *Educational psychologist*, *20*(4), 167-182. https://telearn.hal.science/hal-00190536v1

Shavelson, R. J., Webb, N. M., & Burstein, L. (1986). *Measurement of teaching*. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., pp. 50–91). New York, NY: Macmillan.

Sherman, M. (2014). The role of technology in supporting students' mathematical thinking: Extending the metaphors of amplifier and reorganizer. Contemporary Issues in Technol*ogy and Teacher Education*, *14*(3), 220-246.

Stein, S., & Delaney, V. (2025). Student perceptions vs. practices: AI use in mathematical thinking. In R. M., Zbiek, X. Yao, A. McCloskey, & F. Arbaugh (Eds.), *Proceedings of the 47th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 1686 -1689). Pennsylvania State University. https://doi.org/10.51272/pmena.47.2025

Stein, M. K., & Smith, M. S. (1998). Mathematical tasks as a framework for reflection: From research to practice. *Mathematics teaching in the middle school*, *3*(4), 268-275.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American educational research journal*, *33*(2), 455-488. https://www.jstor.org/stable/1163292

Subramonyam, H., Pea, R., Pondoc, C., Agrawala, M., & Seifert, C. (2024, May). Bridging the gulf of envisioning: Cognitive challenges in prompt based interactions with LLMs. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-19). https://doi.org/10.1145/3613904.3642754

Yoon, H., Hwang, J., Lee, K., Roh, K. H., and Kwon, O. N. (2024). Students' use of generative artificial intelligence for proving mathematical statements. *ZDM–Mathematics Education*, *56*(7), 1531-1551. https://doi.org/10.1007/s11858-024-01629-0