Expert Analysis of AI-Generated Theoretical Analyses in Mathematics Education

Jennifer Kleiman
Cameron Byerley
Kevin C. Moore

Kleiman, J., Byerley, C., & Moore, K. C. (2025). Expert analysis of AI-generated theoretical analyses in mathematics education. 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. 1265-1273). The Pennsylvania State University.

EXPERT ANALYSIS OF AI-GENERATED THEORETICAL ANALYSIS IN MATHEMATICS EDUCATION

Jennifer Kleiman	Cameron Byerley	Kevin C. Moore
University of Georgia	University of Georgia	University of Georgia
jennifer.kleiman@uga.edu	cbyerley@uga.edu	kvcmoore@uga.edu

This study investigates an AI system's strengths and limitations for analyzing theoretical issues in mathematics education. Using a comprehensive exam question about Mathematical Knowledge for Teaching (MKT), we explored how AI, specifically Google's NotebookLM, interprets and distinguishes between Ball's and Thompson's MKT frameworks. Faculty with expertise in MKT research reviewed and identified both surprising competencies and subtle limitations in the AI's theoretical analysis. We present a framework categorizing these limitations as field knowledge, attribution, argumentation, illustration, and artificial neutrality. Our findings suggest that current AI systems can do useful theoretical analysis in math education at a graduate level but demands alertness to nuanced failings which often require expert knowledge to detect. Implications for research and doctoral pedagogy are advanced.

Keywords: Computing and Coding, Data Analysis and Statistics, Mathematical Knowledge for Teaching, Research Methods

As AI systems increasingly demonstrate capabilities for academic analysis across disciplines, we investigate their capacity for theoretical analysis in mathematics education, specifically testing Google's NotebookLM (NLM) on a comprehensive exam question comparing Ball's and Thompson's perspectives on Mathematical Knowledge for Teaching (MKT). While Ball adopted the term MKT from Thompson's 1996 work (Ball et al., 2001), they developed distinct theoretical frameworks reflecting different epistemological stances. Their approaches share common ground but diverge significantly in their theories of knowledge and assessment strategies (Byerley & Thompson, 2017). This makes the analysis a non-trivial academic task, but also one we would engage a novice field member in if interested in MKT. This work contributes to emerging qualitative approaches for evaluating AI-generated academic content, though mathematics education content requires domain-specific expertise and has unique domain-specific challenges.

Using NLM's retrieval-augmented generation (RAG) capabilities, we provided 24 primary sources about MKT and prompted the AI to create a comparative analysis. Two experienced mathematics education researchers then evaluated the output as they would a doctoral student's comprehensive exam response. The following research question guides our qualitative analysis of the AIs writing: "What are the strengths and limitations of an AI-generated theoretical analysis of Ball and Thompson's research on MKT?"

Mathematical Knowledge for Teaching

Thompson's conceptualization of the term knowledge in MKT was heavily influenced by Piaget's work and theories of radical constructivism. His research investigated teachers' underlying conceptual structures for mathematical ideas and teachers' ideas for how those ideas might develop in another (Thompson, 2015). He eventually stopped using the term MKT and started using the term *Mathematical Meanings for Teaching* because he felt the term meaning better captured that he was studying teachers' conceptual structures that were personal to that teacher. Ball et al. (2005) studied MKT by observing elementary teachers in practice to identify the “mathematical knowledge and skill used in the work of teaching” (pp. 16-17). When initially designing an assessment for MKT, Ball thought items measured if teachers knew various facts about students’ mathematical learning. As a result of validation efforts on the assessment, Ball and colleagues’ thinking evolved and they realized they were measuring teachers’ ability to engage in reasoning using both their knowledge of mathematics and students versus the teachers’ factual knowledge (Schilling et al., 2007).

The notable differences between Thompson’s and Ball’s approaches underscore how researcher’s underlying theories of knowledge influence the ways in which they operationalize MKT, and thus the impact their work might have on the preparation and support of teachers.

Retrieval Augmented Generation (RAG), Context Windows, and Academic Work

Anjos et al. (2024) studied AI qualitative analysis of conceptual learning in science education and found two critical technical issues, that of hallucinations and of limited context window. Hallucinations are a well-known issue for Large Language Models (LLMs) where the AI invents information, like citing non-existent research papers (Sun et al., 2024). RAG is designed to reduce hallucinations by directly grounding responses in provided source data. In a recent medical study, NLM achieved 86% accuracy versus just 25% for GPT-4 when analyzing the same reference materials, while providing 95% correct citations (Tozuka et al., 2024). The capacity to curate sources and ground AI responses in selected material creates important control and allows responsible academic use by ensuring that selected sources are reliable. As AI tools become more integrated in research workflows, systems that provide transparent sourcing are preferable to "black box" standard LLMs.

This work contributes to emerging approaches for systematically evaluating AI-generated academic content, including pedagogical frameworks for qualitatively distinguishing AI from human-authored texts (Garib & Coffelt, 2024), multidimensional comparative analysis methods (Berber Sardinha, 2024), and comprehensive evaluation systems for AI-generated educational resources (Huang et al., 2025). We agree with Fonseca et al. (2023) that we are subjectively interpreting the AI output with our own perspectives and biases, AI algorithmic processes have their own implicit and explicit biases, and source artifacts carry their own as well. When using AI generated research analysis it is crucial to interrogate biases in every element of the inference chain and critically reflect on AI outputs to inform subsequent scholarship.

Methodology

We explored AI’s capabilities to answer a comprehensive exam comparing Thompson and Ball’s MKT frameworks, discussing knowledge framing and affordances/constraints of each perspective. After testing four AI systems, we decided that NLM produced the best response.

We employed NLM’s RAG environment with 25 key papers (12 by Ball, 12 by Thompson, plus Tallman 2023) selected based on the second author’s MKT expertise.. We then 1) prompted NLM to respond to the comp question, 2) added resulting essay, back into NLM 3) prompted it to analyze the essay to find any weaknesses, and then 3) prompted it to use its gap analysis to write a better prompt for a new essay. The outcome of that critical reflection cycle was 4) fed back into the AI to generate a new, final version. Materials are available at https://osf.io/57yeu/

Upon first read the NLM essay appeared to us to be of impressively high quality relative to expectations for a comprehensive response during the second year of PhD studies. As a research team, we first discussed our impressions and if the response would pass a comprehensive exam. Then, the second and third authors evaluated the essay rigorously, leaning into their MKT expertise to judge the AI output as if it had been submitted for a comprehensive exam. They considered whether the essay is well-written, answers the question accurately, and demonstrates that the student has adequate knowledge of the subject to operationalize the concept as they plan and conduct a dissertation study in an MKT-related area of study. They did not evaluate if the paper had novel contributions or new theoretical insights that would be worthy of research-journal publication. Because the research team desired to judge the outcome as a comprehensive response, they found it important to make explicit that novel insights were not expected.

We analyzed 74 substantive expert comments using Charmaz and Belgrave's (2019) coding approach through three phases. We generated 72 preliminary codes across the comments, with many comments getting multiple codes. Codes ranged from major (e.g., "cited opinion as fact", “insightful wording”) to minor (e.g., "awkward wording"). We then systematically compared and collapsed the codes into 8 intermediate categories that captured different types of AI limitations and capabilities. Finally, we met as a research team to develop the final five-category framework through consensus discussion, refining categories for theoretical coherence and practical utility. We present our findings below, reviewing the positive then the negative.

AI Competencies and Strengths

Initial expert review concluded that the AI output was comparable to graduate student work, with no clear tells that would identify it as AI-generated. Further, both experts judged the AI to have more accurately discussed literature than many papers submitted for peer-review. The second and third author both thought the paper answered the question asked using the literature provided. The paper accurately summarized many key distinctions between Ball and Thompson’s perspectives and backed it up using appropriate quotes. Neither expert thought a student who submitted this essay would fail the comprehensive exam, but they both had ideas for how this hypothetical student could deepen their thinking in the next stage of their PhD process. The faculty shared uncertainty about whether the AI was fully prepared to operationalize a theory of MKT in its own research because the AI summarized existing literature versus displaying creative analysis and novel examples capturing distinctions between two perspectives on MKT. However, the faculty agreed that this was a common issue with newer human academics. Both ultimately agreed that the response would likely pass comprehensives, with the only exception being a short extension question that prompts the student to provide more concrete examples of the differences.

The research team identified numerous salient attributes that contributed to the assessment of the graduate-level quality of the AI output. The AI analysis showed accurate grasp of MKT literature and effective use of relevant quotes, e.g. this excerpt that is consistent with Thompson’s writing and that was judged to capture key ideas related to Thompson’s thinking:

> Thompson's focus on mathematical meanings provides a deeper, more nuanced understanding of teachers' mathematical thinking. His emphasis on schemes, actions, and reflective abstraction offers a way to understand the dynamic and interconnected nature of mathematical understanding. The focus on meanings also acknowledges the subjective nature of knowledge construction, recognizing that teachers’ actions are often a reflection of their own personal mathematical understandings.

NLM output sometimes exhibited deep historical knowledge of MKT development, e.g.:

“Ball and her colleagues developed a framework that specifies the ‘mathematical tasks of teaching’ which include ‘interpreting and evaluating students’ non-standard mathematical ideas’ (Selling et al., 2016, p. 35). This approach is rooted in Shulman's (1987) notion of pedagogical content knowledge (PCK), expanding and specifying it through empirical investigation.”

The framing, explanation, quote selection, and Shulman connection concisely explains how Ball and colleague's approach builds on prior theory in the field.

The AI used other exemplary quotes from source documents in keen and insightful ways. As an example of a well-chosen quote: "Specialized content knowledge (SCK) is the mathematical knowledge that allows teachers to engage in particular teaching tasks (Hill et al, 2008). For example, a teacher needs SCK to 'explain the meaning of subtraction of a negative number and connect it to moves on the number line in ways that make conceptual sense' (Selling et al., 2016, p. 37)." This quote is accurate and helps illustrate the concept of SCK in a meaningful way.

Later, the essay addresses Thompson's view of knowledge and meaning in another example of incisive commentary. The AI said Thompson "explicitly contrasts his focus on 'meanings' with a focus on 'knowledge,' which he argues is often used as an undefined term", an accurate point as Thompson has frequently pointed out that what it means to "understand topic X" is often left undefined by researchers inside and outside of MKT (Thompson & Saldanha, 2003).

MKT-AI Limitations Framework & Examples

Five general themes of significant limitations emerged from analysis (Table 1). Additional issues of syntax were not theoretically important so were not included in the framework. We share selected quotes that illuminate the issues in each major theme, then discuss implications.

Table 1. MKT-AI Limitations Framework

Category	Description	Examples
Deep Knowledge of the Field	Issues requiring domain expertise to analyze	Characterizing Tallman's work as an alternative to Ball/Thompson rather than building on Thompson
Reference & Attribution	Mishandling source material and citations	Treating opinions stated in sources as established facts
Argumentation	Flow of sentences fails to build an argument	A paragraph addressing differences on culture that ended up not really saying anything
Illustration using Examples	Theoretical descriptions missing concrete examples	Contrasting SCK and Thompson "powerful pedagogical conceptualizations" under the heading "Practical implications" without giving a practical example, only abstract principles
Artificial Neutrality	Imposing or creating consensus	False balance between perspectives Overly nonjudgmental assessments

Deep Field Knowledge

The AI output included passages that sounded reasonable on the surface, but had subtle errors, e.g. in discussion of views of other scholars on MKT: "Alternative perspectives, such as Tallman's Deweyan-based view, propose that PCK must be grounded in an understanding of the functional mechanisms of mathematics learning (Tallman, 2023)." It is certainly true that Tallman references Dewey in his theoretical paper connecting MKT "to its Deweyan roots", but we do not view his work as an alternative to Ball and Thompson's views so much as building upon Thompson's (Tallman, 2023). This category illustrates how the AI occasionally expresses opinions that create incorrect implications, underscoring the need for critical analysis by someone with deep field knowledge to catch subtle errors.

Reference and Attribution

This theme covers issues of pulling from source material: treating an opinion stated in source material as fact, treating a source's description of a third party's work as having been done by the source document author, taking something in a source document at face value when an academic possibly would have dug deeper. Here the AI contrasts Ball and Thompson's views of SCK:

> However, a teacher with a robust mathematical meaning (Thompson) understands *why* students make these errors and how these errors stem from a limited understanding of the meaning of fractions as multiplicative comparisons (Byerley & Thompson, 2017). Such a teacher would be better positioned to design activities and pose questions to shift students' understanding.

This sounds quite authoritative, but Byerley and Thompson (2017) the first sentence is stated as an opinion, not a fact.

Argumentation

This theme covers sections where sentences individually made sense, and were accurate, but failed to make a coherent argument. This was seen at different levels of scale; where one sentence didn't fit well with any of its neighbors, up to where an entire paragraph failed to hold together. These sentences always read well out of context and to a superficial read sounded sensible. Only on critical reflection did these argumentation issues become clear. We caution the reader not to infer that this limitation in the AI output is specific to AI. The research team has confronted this issue in our own and other human writing, particularly by field novices.

The following particularly egregious example is from a section titled "Evolution of Frameworks: From Shulman to Current Developments". Although each sentence is factually correct and grounded in source material, the argument the paragraph presents is unclear at best.

> Considerations of cultural contexts and the needs of English Language Learners (ELLs) add further layers of complexity to the discussion of MKT (Ball et al., 2008). As Thompson notes, teachers' meanings are culturally embedded and difficult to change (Thompson & Milner, 2019). The universality of MKT is an important consideration (Ball et al., 2008). For example, what counts as a valid mathematical explanation or representation may vary across different cultural contexts (Hill et al., 2005). Further, ELLs often bring different mathematical experiences and ways of expressing their understanding. Teachers must be able to draw on their mathematical knowledge and meanings to connect mathematics concepts with students' cultural and linguistic backgrounds.

The AI successfully showed both Ball and Thompson address cultural issues, but the paragraph makes little sense under the "current developments" heading. Importantly, it also fails to sensibly compare or contrast the two views. Is the AI saying that Thompson believes teacher MKT is culturally specific, but Ball considers it universal? If so, it contradicts that in the next sentence.

Similar issues plague the composition elsewhere in spots where the AI-generated sentences read well and make sense taken individually but lack coherence when considered as part of an argument. This structural issue applied mainly to the small scale, where a phrase failed to contribute to a sentence or a sentence to the paragraph, but examples like the one above show that even an entire paragraph could be less than coherent.

Failure to Illustrate

Many of the AI paragraphs were both accurate and well-written, such as the following:

> Thompson's framework suggests that professional development must focus on transforming teachers' underlying mathematical meanings and ways of thinking (Thompson, 2015). This implies going beyond teaching specific techniques and engaging teachers in activities that challenge and expand their own conceptualizations of mathematics. Thompson suggests that professional development should focus on helping teachers develop ‘powerful pedagogical conceptualizations’ (Silverman & Thompson, 2005) of the mathematics they teach.

Despite its accuracy and cohesiveness, the paragraph is a perfect exemplar of a failure to illustrate. The reader is told that Thompson’s views imply engaging teachers in professional development to deepen their conceptual understandings, which naturally invites the reader to wonder what such a professional development would look like. Specific suggestions for professional development are discussed in Byerley & Thompson (2017) so the AI had access to that information, but did not use it here.

The AI often gave a good thousand-foot analysis but did not always provide examples of critical epistemological differences that drive Ball and Thompson's distinct approaches. For instance, when discussing Thompson’s 'powerful pedagogical conceptualizations' versus Ball's SCK, the AI used the correct terms but did not illuminate with specific examples how these concepts reflect fundamentally different views of teacher knowledge.

In essence, its key failings were not blatant errors but could be seen as a failure of empathy. An expert might have understood that certain concepts would raise questions in a reader that invited clarification. In many of those situations, but not all, the AI did not employ concrete examples that could have illuminated the distinctions it was making to the reader. Without grounding theoretical distinctions in specific examples, the analysis missed opportunities to help readers construct actionable understanding of these contrasting frameworks. Taken as a comp exam response, while it stated differences between the two perspectives, the AI writer did not illustrate that they grasped the implications of those. That understanding would be critical to operationalize the perspectives during the design and execution of a dissertation study.

In other places, the AI essay did allude to examples but did not give enough details of the examples to be coherent or illustrate perspective differences. For instance:

> Another example, given by Thompson (2016, p. 463), is that a teacher with a strong understanding of variable will understand that the solutions to an equation are the values that make the equation true, and can help students see that connection to a function's domain and image.

This is accurate, relevant, and coherently contributes to the argument being made in that section. However, there is a good deal of background theory that is needed to understand why Thompsons says it is significant that a solution is a (static) value that makes an equation true, as opposed to a dynamic variable that is part of a functional relationship.

Artificial Neutrality

AI chatbots have been shown to create false consensus (Choi et al., 2025). The AI essay concludes with a clear example of this, claiming that “By integrating these [Ball and Thompson’s] perspectives, mathematics education can move towards more effective models of teacher education and professional development, thereby ultimately improving student learning.” Novice human math educators may share this tendency to want to combine and unite conceptual frameworks, but a more sophisticated understanding of theory would take compatibility, coherence, and utility into account before proposing to mash ideas together. Whether this counts as a technical AI error or not is debatable, but we judged it to be a weak point of the AI essay.

AI Capabilities for Mathematics Education Research: Lessons Learned

Our MKT-AI Limitations Framework identified five themes, patterns in how AI weaknesses in grappling with theory can manifest. The framework's major categories of Deep Field Knowledge, Reference & Attribution, Argumentation, Failure to Illustrate, and Artificial Neutrality helped us become alert to subtle failings and led us to become more aware of these issues manifesting elsewhere in the text. NLM clearly succeeded in both small and large ways and demonstrated impressive capabilities to synthesize a large body of knowledge, identify important theoretical constructs, use them in a mostly-coherent narrative, and employ correct citations and references. However, its failings were evident to experts upon close reading and critical reflection. Some of these issues may be addressed by further prompting the AI in ways that take these themes into account, suggesting direction for another critical study.

Pedagogical Implications

An AI system that can produce a "passing" comprehensive exam response could be an opportunity to leverage AI capabilities in doctoral training. For instance, AI-generated theoretical analyses could serve as discussion prompts in doctoral seminars, with students practicing critical evaluation of arguments and identifying subtle flaws or missing insights. This develops their own expertise while familiarizing them with AI's strengths and limitations. Doctoral students are still required to present oral defense so demonstration of mastery of their material without AI scaffolding is still an inherent part of the process.

Gaps between AI's shallow-seeming expertise and deep scholarly understanding exemplify aspects of doctoral training we might emphasize more explicitly. While the AI could report what scholars said, it struggled to demonstrate judgment that anticipates reader needs or selects illuminating examples—skills we want doctoral students to develop but may not explicitly teach. AI responses could provide concrete examples for discussing higher-level scholarly capabilities.

Additionally, AI tools might serve as "practice partners" for comprehensive exam preparation. Students could critique AI-generated responses to practice questions, developing their analytical skills while gaining confidence in their own expertise. This also raises questions about how we assess doctoral student understanding. With the progress of AI and increasing presence as part of academic workflow, perhaps we need to more explicitly focus on skills that center human scholarly judgment used in critical partnership with AI.

Anticipating the Impact of Future AI Developments

Looking ahead, advances in AI will likely enhance all aspects of mathematics education research. Future developments will possibly make our concerns obsolete, but our analysis suggests present practice should focus on specific applications where AI can support research in a reliable or at least predictable fashion. Literature search, research synthesis, data analysis, and even hypothesis generation can be productive activities if they are paired with expert oversight. Our methodology contributes to emerging foundations for evaluating AI-generated theoretical analysis, joining concurrent methodological developments across disciplines in establishing systematic approaches for AI-academic collaboration.

Could iterative cycles of generation, criticism, and prompt refinement converge on a stable, high-quality theoretical analysis? Our experience suggests that one such cycle improved the AI output by addressing identified gaps and omissions. The next step is to test a multi-iteration approach systematically using newer reasoning models, with prompts explicitly designed to target the error patterns we identified. Now and in future AI-human research collaboration, human experts will be needed to evaluate each step critically.

Zbiek, R. M., Yao, X., McCloskey, A., & Arbaugh, F. (Eds.). (2025). *Proceedings of the forty-seventh annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Pennsylvania State University. https://doi.org/ 10.51272/pmena.47.2025

References

Anjos, J. R. dos, Souza, M. G. de, Neto, A. S. de A., & Souza, B. C. de. (2024). An analysis of the generative AI use as analyst in qualitative research in science education. *Revista Pesquisa Qualitativa*, *12*(30), Article 30. https://doi.org/10.33361/RPQ.2024.v.12.n.30.724

Ball, D. L., Hill, H. C., & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator*, *29*(1), 14-17, 20-22, 43-46. http://hdl.handle.net/2027.42/65072

Ball, D. L., Lubienski, S. T., & Spangler, D. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (Vol. 4, pp. 433–456). American Educational Research Association.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*(5), 389–407. https://doi.org/10.1177/0022487108324554

Berber Sardinha, T. (2024). AI-generated vs human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics*, *4*(1), 100083. https://doi.org/10.1016/j.acorp.2023.100083

Byerley, C., & Thompson, P. W. (2017). Secondary mathematics teachers' meanings for measure, slope, and rate of change. *The Journal of Mathematical Behavior*, *48*, 168–193. https://doi.org/10.1016/j.jmathb.2017.09.003

Charmaz, K., & Belgrave, L. L. (2019). Thinking about data with grounded theory. *Qualitative Inquiry*, *25*(8), 743–753. https://doi.org/10.1177/1077800418809455

Choi, J., Hong, Y., & Kim, B. (2025). *People will agree what I think: Investigating LLM's false consensus effect* (No. arXiv:2407.12007). arXiv. https://doi.org/10.48550/arXiv.2407.12007

Fonseca, A. L. A. D., Chimenti, P. C. P. D. S., & Suarez, M. C. (2023). Using deep learning language models as scaffolding tools in interpretive research. *Revista de Administração Contemporânea*, *27*(3), e230021. https://doi.org/10.1590/1982-7849rac2023230021.en

Garib, A., & Coffelt, T. A. (2024). DETECTing the anomalies: Exploring implications of qualitative research in identifying AI-generated text for AI-assisted composition instruction. *Computers and Composition*, *73*, 102869. https://doi.org/10.1016/j.compcom.2024.102869

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, *39*(4), 372–400. https://doi.org/10.5951/jresematheduc.39.4.0372

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371–406.

Hoover, M., Mosvold, R., Ball, D. L., & Lai, Y. (2016). Making progress on mathematical knowledge for teaching. *The Mathematics Enthusiast*, *13*(1–2), 3–34. https://doi.org/10.54870/1551-3440.1363

Huang, Q., Lv, C., Lu, L., & Tu, S. (2025). Evaluating the quality of AI-generated digital educational resources for university teaching and learning. *Systems*, *13*(3), 174. https://doi.org/10.3390/systems13030174

Schilling, S. G., Blunk, M., & Hill, H. C. (2007). Test validation and the MKT measures: Generalizations and conclusions. *Measurement: Interdisciplinary Research and Perspectives*, *5*(2–3), 118–128. https://doi.org/10.1080/15366360701487146

Selling, S. K., Garcia, N., & Ball, D. L. (2016). What does it take to develop assessments of mathematical knowledge for teaching? Unpacking the mathematical work of teaching. *The Mathematics Enthusiast*, *13*(1–2), 35–51. https://doi.org/10.54870/1551-3440.1364

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Silverman, J., & Thompson, P. W. (2005). Investigating the relationship between mathematical understanding and teaching mathematics. In S. Wilson (Ed.), *Proceedings of the twenty-seventh annual meeting of the International Group for the Psychology of Mathematics Education*. Virginia Tech.

Sun, Y., Sheng, D., Zhou, Z., & Wu, Y. (2024). AI hallucination: Towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanities and Social Sciences Communications*, *11*(1), 1–14. https://doi.org/10.1057/s41599-024-03811-x

Tallman, M. A. (2023). What makes pedagogical content knowledge "pedagogical"? Reconnecting PCK to its Deweyan foundations. *The Mathematics Educator*, *31*(1).

Thompson, A. G., & Thompson, P. W. (1996). Talking about rates conceptually, Part II: Mathematical knowledge for teaching. *Journal for research in Mathematics Education*, *27*(1), 2-24.

Thompson, P. W. (2015). Researching mathematical meanings for teaching 1, 2. In L. D. English & D. Kirshner (Eds.), *Handbook of international research in mathematics education* (pp. 435-461). Routledge.

Thompson, P. W., & Milner, F. (2019). Teachers' meanings for function and function notation in South Korea and the United States. In H.-G. Weigand, W. McCallum, M. Menghini, M. Neubrand, & G. Schubring (Eds.), *The Legacy of Felix Klein* (pp. 55–66). Springer International Publishing. https://doi.org/10.1007/978-3-319-99386-7_4

Thompson, P. W., & Saldanha, L. A. (2003). Fractions with multiplicative reasoning. In J. Kilpatrick, G. Martin, & D. Schifter (Eds.), *Research companion to the Principles and Standards for School Mathematics* (pp. 95-114). National Council of Teachers of Mathematics.

Tozuka, R., Johno, H., Amakawa, A., Sato, J., Muto, M., Seki, S., Komaba, A, & Onishi, H. (2024) Application of NotebookLM, a large language model with retrieval-augmented generation, for lung cancer staging. *Japanese Journal of Radiology, 43*, 706–712. https://doi.org/10.1007/s11604-024-01705-1