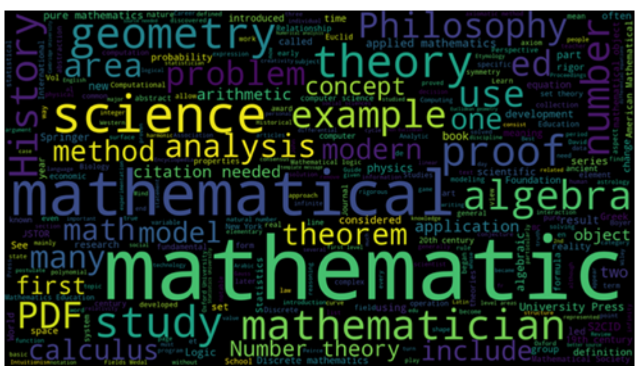
Proceedings for the 52nd Annual Meeting of the Research Council on Mathematics Learning

Celebrating 52 years of Research on Mathematics Learning



Innovating and Integrating: Advancing Mathematics Learning Across Disciplines

March 6 – 8, 2025
College Station, Texas

RCML Officers & Conference Team

THANK YOU TO OUR REVIEWERS

For citation of articles in the Proceedings:

Authors. (2025). Title of paper. In A. Wheeler, & P. Kaur Bharaj (Eds.), *Proceedings of the 52nd Annual Meeting of the Research Council on Mathematics Learning.* (pp.XX- XX). RCML.

Editorial Assistance:

Reuben S. Asempapa, Penn State University Harrisburg, Middleton, PA.

Graduate Student Editorial Assistant:

Margeaux Smith, Baylor University, Waco, TX

Publication Acceptance Rate:

Accepted 14 manuscripts out of 24 submissions. Acceptance Rate of 58% (last year 54%).

RCML History

The Research Council on Mathematics Learning, formerly The Research Council for Diagnostic and Prescriptive Mathematics, grew from a seed planted at a 1974 national conference held at Kent State University. A need for an informational sharing structure in diagnostic, prescriptive, and remedial mathematics was identified by James W. Heddens. A group of invited professional educators convened to explore, discuss, and exchange ideas especially in regard to pupils having difficulty in learning mathematics. It was noted that there was considerable fragmentation and repetition of effort in research on learning deficiencies at all levels of student mathematical development. The discussions centered on how individuals could pool their talents, resources, and research efforts to help develop a body of knowledge. The intent was for teams of researchers to work together in collaborative research focused on solving student difficulties encountered in learning mathematics.

Specific areas identified were:

1. Synthesize innovative approaches.
2. Create insightful diagnostic instruments.
3. Create diagnostic techniques.
4. Develop new and interesting materials.
5. Examine research reporting strategies.

As a professional organization, the **Research Council on Mathematics Learning (RCML)** may be thought of as a vehicle to be used by its membership to accomplish specific goals. There is an opportunity for everyone to actively participate in **RCML**. Indeed, such participation is mandatory if **RCML** is to continue to provide a forum for exploration, examination, and professional growth for mathematics educators at all levels.

The Founding Members of the Council are those individuals that presented papers at one of the first three National Remedial Mathematics Conferences held at Kent State University in 1974, 1975, and 1976.

Table of Contents

Innovating and Integrating: Mathematical Learning in K-12 Education

MISSION TO MARS: A CURRICULUM EQUITY ANALYSIS

Stefanie D. Livers
Bowling Green State University
slivers@bgsu.edu

Erin Rich
Indiana University
richee@iu.edu

Chloe S. Bolyard
Missouri State University
cbolyard@missouristate.edu

Ashleigh R. Bettum
Bowling Green State University
abettum@bgsu.edu

Jennice McCafferty
Missouri State University
jmwright@missouristate.edu

Using the Equity-Oriented STEM Literacy Framework, we analyzed NASA's Mars Curriculum within a curriculum criticism study through a deductive coding process. Curriculum criticism as a form of analysis allowed for an equity-oriented assessment. We identified the strengths and the areas that need attention to foster empathy, STEM identity development, and positive dispositions. Implications include ensuring that curriculum activities provide opportunities for students to see themselves, learn through perseverance, and foster critical thinking to solve authentic problems.

A rapidly changing climate, escalating social inequalities, and the relentless pace of technological change are among the many challenges facing society. As today's kindergartners step into their high school graduation gowns, they are likely to encounter an even more complex world. Predictions suggest more extreme climate events, widening social rifts, and transformative technological innovations, particularly in automation and artificial intelligence. Mohr-Schroder and colleagues (2020) contend that these challenges make STEM education necessary to prepare students to be effective problem solvers and critical thinkers through real world learning experiences. Moreover, they feel students not only need to show competency in content areas but also connect them to solve complex problems.

The norm in school instruction is isolated teaching of core subjects even though real world problem solving is an integrated application of multiple content knowledge bases (Honey et al., 2014). In comparison to STEM instruction that is thought of as "working in the context of complex phenomena or situations on tasks that require students to use knowledge and skills from multiple disciplines" (p. 52). Because of this, STEM integration is promoted because solutions to authentic problems, such as economic, environmental, and social, are often interdisciplinary (English et al., 2016; Honey et al., 2014).

Marginalized students (e.g., BIPOC students, multilingual learners, and students with disabilities) often experience impoverished teaching that is riddled with skill exercises and procedures that do not promote higher order thinking (Delpit, 2012). Additionally, many STEM experiences lack connections and engagement of marginalized students. The lack of presence of diverse students in STEM education and disconnect from their experiences (Calabrese Barton & Tan, 2019) is problematic if we are seeking to prepare students for this ever-changing world. Students' mathematics scores are improved when students are engaged in integrated learning experiences (Gadanidis et al., 2017**;** Gadanidis & Hughes, 2011). When we start young children in early STEM-based instruction, they have increased reading, writing, and mathematics scores, with evidence of gains for marginalized students (Sarama et al., 2018).

Curricula outline students' learning opportunities (Milner, 2010). As we work to ensure high-quality STEM opportunities for all students, curriculum analysis is necessary to ensure students' opportunities and access to learn. Gao et al. (2020) conducted a systematic literature review of 49 articles regarding STEM curriculum and found overall that materials need to be more explicit and systematic. They argue that how content in STEM curriculum is intersected, presented, and assessed needs to be clearly developed and laid out. Moreover, the learning process and supporting practices should be evident from congruent objectives and assessments highlighting the complexity of integrated STEM.

To disrupt systems of oppression and privilege, it is important to ensure that criteria to support empathy, empowerment, dispositions, identity development, utility and applicability, critical thinking and problem solving are present in STEM curricula (Jackson et al., 2021) so students will be confident to be future change agents. By examining what is omitted from the curriculum, educators and students can identify and challenge biases, stereotypes, and power dynamics that limit opportunities for marginalized groups (Milner, 2010). For this curriculum analysis, we chose NASA's Mars curricula. In a literature review of studies using Mars curricula, none focused on equity. Most studies investigated implications on implementing Mars curricula, showing positive learning effects (Pilla et al., 2021; Raghavan et al., 1998). Salmi et al. (2023) found that a Mars learning intervention had a positive impact on 11–13-year-olds.

Purpose

The purpose of our study was to examine NASA's Mars curriculum with the Equity-Oriented STEM Literacy Framework to assess how it addresses the ability to produce societal change agents. More specifically, our study is guided by the following research question: *What STEM equity-oriented strengths and opportunities exist within a NASA curricular unit about Mars*?

Theoretical Framing

We chose the Equity-Oriented STEM Literacy Framework to ground this study. Given the historical trends that STEM education should be accessible to all students, we chose a framing that promotes equity and access for all. The Equity-Oriented STEM Literacy Framework has six criteria that work together to provide experiences that prepare students to be societal change agents. The six criteria are the following: *empowerment; empathy; dispositions; identity development; utility and applicability*; and *critical thinking and problem solving*. To disrupt systems of oppression and privilege, it is important to ensure that these criteria are present in STEM curricula (Jackson et al., 2021). By examining what is omitted from the curriculum, educators and students can identify and challenge biases, stereotypes, and power dynamics that limit opportunities for marginalized groups (Milner, 2010).

Methodology

This qualitative study incorporated a curriculum criticism approach. Curriculum criticism (Sherman & Webb, 1988) is an inquiry evaluation method that incorporates inclusion of appraisal to analyze curriculum with a critical perspective. Curriculum criticism reveals more than intended outcomes including deeper understandings and perspectives. The goal of critical criticism is "to reveal and explain the meaning and complexity" of the unit of analysis (p. 163). For this study, we sought to reveal if the chosen curriculum was equity oriented.

Data Source

We chose to analyze a series of STEM lessons from NASA's Jet Propulsion Laboratory entitled *Mission to Mars* (NASA, n.d.). The standards-aligned unit developed by NASA guides students through a series of guided lessons to learn more about and plan a mission to Mars. The unit is composed of 90 resources broken into seven lessons: *Learn About Mars, Plan Your Mission, Design Your Spacecraft, Launch Your Mission, Land on Mars, Surface Operations*, and *Sample Handling*. Each of the seven lessons contains resources for educators, including educator guides, student projects, student articles, and student videos. While the focus of our coding was the educator guides, some guides linked to additional resources for students.

Analysis

Our deductive coding involved multiple rounds of analysis. After each round of coding, we met as a research team to discuss the coding process. Round One allowed us to narrow the scope of the data source, choosing to focus on NASA's *Mission to Mars* unit and coding instances for concepts from the Equity-Oriented STEM Literacy Framework. During Round Two, the team individually coded 20 portions from the data source and compared codes using Hypothesis (n.d.), a free social annotation browser extension. We created a group in Hypothesis that allowed us to collect each team member's codes. Then, the team came to a consensus regarding the meaning of each of the criteria and our coding. In Round Three, we coded the rest of the *Mission to Mars* lessons separately and then created a table to come to a consensus on coding; if two or more of us aligned, the third coder broke the tie. We organized our data around codes from the Equity-Oriented STEM Literacy Framework, with the coded text, and location for all codes with at least two coders coding the element from the educator's guide the same (see Table 1).

Table 1

Steamy Mars Annotations

Code	Example	Location
Empathy	"Try this STEM strategy card game to get students thinking like the NASA scientists and engineers working on these exciting missions to the Moon, Mars, and beyond, as they prepare to join the Artemis Generation."	Educator Guide - NASA Space Voyagers: The Game
STEM Identity Development	"This project will help students understand the engineering process by allowing them to design a robotic insect for an extraterrestrial environment"	Educator Guide - Design a Robotic Insect
Critical Thinking and Problem Solving	"The difficulty can be further increased by having students consider the movement of planets as they orbit the Sun. How can relays be placed to ensure information can be transmitted regardless of where two planets are relative to each other and the Sun.	Educator Guide - Build a Relay Inspired by Space Communications

Findings

In our curriculum analysis of *Mission to Mars*, we sought to address the following question: *What STEM equity-oriented strengths and opportunities exist within a NASA curricular unit on*

Mars? Using the Equity-Oriented STEM Literacy Framework, we identified numerous strengths and opportunities present in the unit related to the six concepts in the framework: *empathy; dispositions; STEM identity development; empowerment; critical thinking and problem solving;* and *utility and applicability*. On the surface, it might appear that each lesson as a whole, addresses all of these criteria, but upon closer examination, we found that only three criteria (*critical thinking and problem solving, utility and applicability*, and *empowerment*) were present in the majority of the Educator Guides. Additionally, two criteria (*empathy* and *STEM identity development*) were present in about a quarter or fewer of the educator guides.

The strengths we identified within *Mission to Mars* rest mostly in its promotion of *critical thinking and problem solving, empowerment*, and *utility and applicability*. In approximately 80 percent of its Educator Guides, the unit promoted *critical thinking and problem solving*, providing students with many opportunities to apply their critical thinking skills to solve complex problems. *Mission to Mars* also promoted *empowerment* in approximately 80 percent of its Educator Guides, frequently giving students opportunities to collaborate with peers, have choice in how they approach content and solutions, and see role models from similar backgrounds. *Utility and applicability* were present in over two-thirds of the lessons, providing relevance for the content in its connection to space exploration and scientists' experiences. This was often identified within the Background section of Educator Guides and provided contextual information to students about how the content of the lesson connects to NASA scientists.

We also identified areas where *Mission to Mars* could be improved more thoroughly and consistently promote *empathy, STEM identity development*, and *dispositions*. Only nine percent of the Educator Guides provided students with an opportunity to connect with problems on a personal level and to understand the impact on others. Enhancing this aspect of *Mission to Mars* could provide transformative learning experiences that serve as a potential bridge for students who have otherwise encountered barriers to STEM (Jackson et al., 2021). Moreover, *STEM identity development*, which enables students to access the material from their cultural and linguistic position, connecting STEM content to their community and identity, was present in only 12 percent of *Mission to Mars*. Finally, *Mission to Mars* promoted productive *STEM dispositions* giving students opportunities to engage in hands-on, trial-and-error investigations in approximately half of its Educator Guides. Such dispositions are critical for demonstrating to students that STEM is a field for everyone (Jackson et al., 2021).

Discussion and Implications

STEM education aims to equip students with the necessary tools and mindsets to become active and engaged citizens who can leverage STEM learning to make a positive impact on their communities and the world. When students feel empowered and see their experiences reflected in the curriculum, learning becomes more inclusive and accessible for everyone. Developing and/or analyzing curricular materials through an equity lens, such as the Equity-Oriented STEM Literacy Framework, enables educators to ensure they are promoting an inclusive learning experience for all students. In STEM education, this sort of curricular criticism or audit is crucial to ensure that content goes beyond promoting critical thinking and problem solving. It should personally connect with students, to affirm their culture and identity, to remove barriers to STEM, and to empower them with the knowledge, skills, and confidence to use STEM to make a positive impact on *their* communities and the world.

Based on our analysis of *Mission to Mars*, we identified potential areas for greater attention and emphasis in STEM curricula, specifically ways in which materials promote *empathy, STEM identity development*, and productive STEM *dispositions*. For example, we feel that the *Mission to Mars* curriculum should include more opportunities for students to see themselves in STEM fields through more real-life examples of individuals working in these fields. Additionally, we suggest that the curriculum promotes productive STEM *dispositions* by including more activities that require trial and error and perseverance by students. In terms of these three criteria, the Mission to Mars curriculum needs to be more explicit and systematic (Gao et al., 2020).

References

Calabrese Barton, A., & Tan, E. (2019). Designing for rightful presence in STEM: The role of making present practices. *Journal of the Learning Sciences*, *28*(4-5), 616–658.

Delpit, L. (2012). *Multiplication is for white people: Raising expectations for other people's children*. The New Press.

English, L. D., King, D. & Smeed, J. (2016). Advancing integrated STEM learning through engineering design: Sixth-grader students' design and construction of earthquake resistant buildings. *Journal of Educational Research*, *110*(3), 255–271.

Gao, X., Li, P., Shen, J., & Sun, H. (2020). Reviewing assessment of student learning in interdisciplinary STEM education. *International Journal of STEM Education*, *7*, 1–14.

Gadanidis, G. & Hughes, J. (2011). Performing big math ideas across the grades. *Teaching Children Mathematics, 17*(8), 486–496.

Gadanidis, G., Hughes, J. M., Minniti, L., & White, B. J. (2017). Computational thinking, grade 1 students and the binomial theorem. *Digital Experiences in Mathematics Education*, *3*, 77-96.
Honey, M., Pearson, G. & Schweingruber (Eds.). (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research*. National Academies Press.
Hypothesis. (n.d.). Hypothesis: An open-source annotation tool. https://web.hypothes.is/
Jackson, C., Mohr-Schroeder, M. J., Bush, S. B., Maiorca, C., Roberts, T., Yost, C., & Fowler, A. (2021). Equity-oriented conceptual framework for K-12 STEM literacy. *International Journal of STEM Education*, *8*, 1–16. https://doi.org/10.1186/s40594-021-00294-z
Milner IV, H. R. (2010). Culture, curriculum, and identity in education. In *Culture, curriculum, and identity in education* (pp. 1–11). Palgrave Macmillan.
Mohr-Schroeder, M. J., Bush, S. B., Maiorca, C., & Nickels, M. (2020). Moving toward an equity-based approach for STEM literacy. In *Handbook of research on STEM education* (pp. 29–38). Routledge.
NASA Jet Propulsion Laboratory (n.d.). *Mission to Mars unit*. https://www.jpl.nasa.gov/edu/teach/activity/mission-to-mars-unit/
Pilla, E., Salmi, H., & Thuneberg, H. (2021). STEAM-Learning to Mars: Students' ideas of space research. *Education Sciences*, 11(3), 122-142. https://doi.org/10.3390/educsci11030122
Raghavan, K., Sartoris, M. L., & Glaser, R. (1998). Why does it go up? The impact of the MARS curriculum as revealed through changes in student explanations of a helium balloon. *Journal of Research in Science Teaching*, 35(5), 547-567. https://doi.org/10.1002/(SICI)1098-2736(199805)35:5%3C547::AID-TEA5%3E3.0.CO;2-P
Salmi H., Thuneberg, H., & Bogner, F. (2023). Is there deep learning on Mars? STEAM education in an inquiry-based out-of-school setting. *Interactive Learning Environments*, 31(2), 1173-1185, https://doi.org/10.1080/10494820.2020.1823856
Sarama, J., Clements, D., Nielsen, N., Blanton, M., Romance, N., Hoover, M Staudt, C., Baroody, A., McWayne, C., & McCulloch, C. (2018). Considerations for STEM Education from PreK through Grade 3. *Community for Advancing Discovery Research in Education (CADRE)*.
Sherman, R. R., & Webb, R. B. (1988). *Qualitative Research in Education: Focus and Methods. The Falmer Press*.

FIFTH GRADE STUDENTS' MULTIPLICATIVE REASONING WHEN ENGAGING WITH PARTIAL PRODUCTS

Sanampreet Gill	Karl W. Kosko	Temitope Egbedeyi
Kent State University	Kent State University	Kent State University
sgill13@kent.edu	kkosko1@kent.edu	tegbedey@kent.edu

Despite the importance of multiplicative reasoning (MR) in mathematics, there is limited empirical study of the role of MR in multi-digit multiplication tasks. This study examined children's multiplicative reasoning within their use of partial products. This mixed methods study of 16 fifth-grade students revealed that half of participants did not attend to partial products, and those that did tended to operate with the second or third multiplicative concept.

Introduction

Although there is significant scholarship on multi-digit addition and subtraction, research on multi-digit multiplication has, historically, been limited in scope and magnitude (Fuson, 2003; Harrison, 2013; Hickendorff et al., 2019). Prior work has typically either focused on the use of symbolic strategies and associated reasoning (Hickendorff et al., 2019), or the relationship between students visual and symbolic approaches to multi-digit multiplication (Izsàk, 2005; Larsson et al., 2017). In particular, the role of partial products in multiplication is often noted, but analysis of how children reason about them is largely absent from the literature (Fuson, 2003; Iszák, 2005). This study emerged as a contribution to the existing gap to gain more insight into students' multiplicative reasoning and conceptual knowledge, with the purpose being to examine how children's multiplicative reasoning relates to how they solve multi-digit multiplication tasks.

Theoretical Framework & Literature Review

The current paper uses Scheme Theory to examine children's multiplicative reasoning, with particular focus on the multiplicative concepts as described by Hackenberg (2010). Prior to multiplicative reasoning, students demonstrate pre-multiplicative reasoning by using a standard algorithm or counting by 1s to solve the multiplication task. For instance, a child may use a standard algorithm to solve 4x19. However, when "carrying the 3," they may not be able to explain that it represents 30. If pressed to represent 4x30 or 4x3 visually, they will be unable to do so. At the first multiplicative concept (MC1), a child is anticipated to use any strategy like skip-counting, repeated addition, and doubling to solve the given task. with a limited ability to visualize part of the partial products. For example, a child may solve 4x19 successfully by doubling 19 (or 19+19) and then doubling 38 to get 76, but when "carrying the 3," they may not be able to explain that it represents 30. If pressed to represent 4x30 visually, they may do it with some difficulty (but will easily be able to do so for 4x3).

At the second multiplicative concept (MC 2), a child is anticipated to solve and represent four partial products with an area model (arrays rods or base 10 blocks) but explain it conceptually. For example, a child may be able to solve 19x24 successfully with a standard algorithm but unable to represent all the partial products on a dot paper (i.e., visually shows 10x20 and 9x4, but not the other two partial products). When a child can solve and represent all the partial products with an area model, dot paper, or other manipulative, such a child is considered to be at the third multiplicative concept; during this stage, the child is anticipated to solve and interpret four or more partial products/quotients. For instance, a child solves 19x24 successfully with a standard algorithm and represents it visually on dot paper, area model, or base 10 blocks (i.e., visually shows 10x20, 10x4, 9x20, and 9x4).

As noted earlier, examination of children's reasoning with multi-digit multiplication has been limited in scope and magnitude (Fuson, 2003; Harrison, 2013; Hickendorff et al., 2019). A clear indication from much of the research on students' symbolic approaches to multiplication suggests that good recall of multiplication facts and accuracy in answers is an indicator of success (Javornik & Lipovec, 2020; Lin & Kubina, 2005). However, many students are unable to meaningfully connect their symbolic algorithmic work with a drawn representation (Hurst & Hurrell, 2016; Larsson et al., 2017). Indeed, Larsson et al. (2017) found that, when pressed for explanation, such students relied on repeated addition instead of any multiplicative reasoning. Examining how children used bundles of sticks when modeling multi-digit multiplication, Hurst and Hurrell (2016) found that although all students "seemed to have a robust recall of multiplication facts…this facility to recall multiplication facts was not an indicator that the students had a conceptual understanding of multiplication" (p. 37).

Key in understanding children's skill and reasoning with multi-digit multiplication is their developed understanding of partial products (Hickendorff et al., 2019). Most scholars examining students' conceptual understanding of partial products have advocated introducing it with arrays or some other form of area model to allow for visualizing the decomposition of numbers into their base-ten components (Izsàk, 2005; Young-Loveridge & Mills, 2009). In examining this scholarship, it appears that students may initially learn to solve problems with two partial products (i.e., $15\times24 = 15\times20 + 15\times4$) before learning to do so with four or more partial products (i.e., 15×24 = 10×20 + 10×4 + 5×20 + 5×4). Despite this assumption in the literature, there is little empirical study of how this aligns with different levels of multiplicative reasoning. Such is the central focus in the current paper: to examine the relationship between children's multiplicative reasoning and their strategies in using partial products.

Methods

Sample & Procedure

Participants included a convenience sample of 16 fifth-grade students. Participants completed the Multiplicative Reasoning Assessment (MRA). The MRA is a 21-item assessment of children's multiplicative concepts. Validity evidence includes psychometric data for internal structure and test content, and cognitive interviews for response processes (Kosko, 2019). Raw dichotomous responses on the MRA were transformed via a Rasch model into a continuous scale. Four students, representing each of the multiplicative reasoning levels assessed by the MRA (pre-multiplicative, MC1, MC2, & MC3) were purposefully assigned within one of each of four conditions. Each condition provided a specific visual: base-ten blocks, array rods, or dot array paper as a visual for explaining their multiplication (see Figure 1). Visualizations were modeled with the task 2×12, which was considered accessible to students not yet able to skip-count given the connection to doubling. Next, participants were asked to solve four multiplication tasks, with follow-up questions for assessing their reasoning, use of the visual, and any written algorithms used: 4×19, 15×24, 8×33, and 13×27.

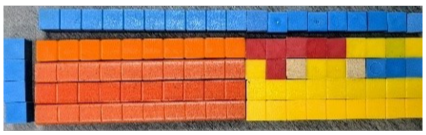
During each interview, participants wore a Pupil Core eye-tracking headset, which recorded their pupil-gaze and front view. Eye tracking is an exploratory method used to analyze students' behavior through gaze patterns, which allows researchers to capture real-time data on how students engage with tasks. It can distinguish gaze patterns, pupil dilation, and fixations (Duchowski, 2007). In general, numerous but shorter duration fixations tend to indicate lower working memory load and higher ability whereas fewer but longer duration fixations indicate more demands on working memory (Rayner, 2006). However, where such fixations matters as much as their duration and number. In the present study, we examined pupil-gaze qualitative in concert with other video-based interview data. Given the exploratory nature of our analysis, this seemed appropriate.

A data-transformation variant of convergent mixed methods design was used in the current study (Creswell & Plano Clark, 2018). We collected interview data of students with eye-tracking glasses and examined their spoken explanations, hand gestures, eye-gaze, use of manipulatives, and written mathematics. This multi-modal data allowed for a richer analysis of the meaning conveyed by students' actions. Following qualitative analysis, emergent themes were transformed into ordinal data and used in a correlational analysis with MRA scores.

Figure 1

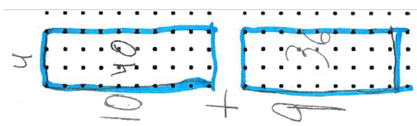
Visualizations Used in Interviews

Base Ten Blocks



Array Rods

Dot Array Paper



Array Rods +

Analysis & Results

Qualitative Analysis & Findings

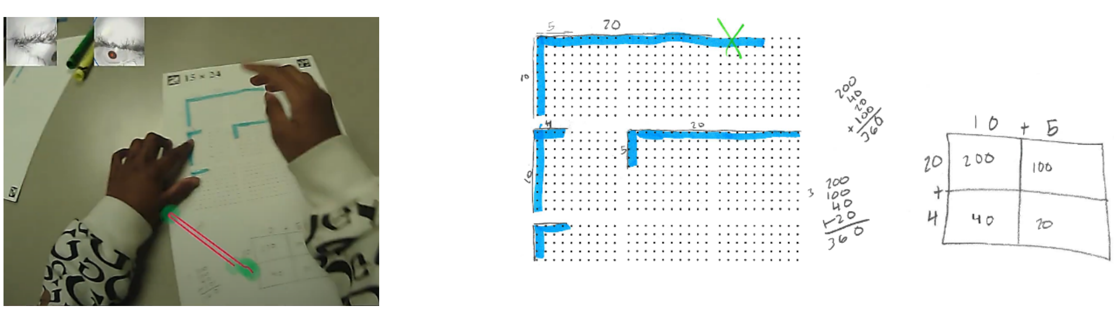
We engaged in three rounds of qualitative analysis across all three authors, focusing on how students conveyed multiplicative meaning. Students' written work and eye-tracking videos were used to examine gestures, eye-gaze, use of manipulatives, verbal descriptions and written mathematics. After an initial round of individual open coding, authors met and examined emergent themes for overlap. A second round of analysis confirmed three emergent themes for analysis: students did not attend to partial products; students attended to no more than two partial products; students attended to more than two partial products. We used these themes to examine students' work with particular tasks in the next round of analysis, as some students did not consistently convey their use of partial products. In the paragraphs below, we briefly summarize these themes with three cases with particular attention to students' solving 15×24.

Did Not Attend to Partial Products

Half of participants were unable to attend to partial products. Many demonstrated either this by using additive reasoning or doubling approaches, but such evidence often did not become apparent until they were asked to explain or visually represent their reasoning. Among these individuals, Amanda was of particular interest because she could easily solve any problem, so long as it was represented symbolically. Amanda was able to quickly able to use the box method to find a solution to 15×24 before adding each multiplicand. Yet, when asked to demonstrate her reasoning with arrays, Amanda began counting ten dots and five dots to draw two lines in the top-left corner of the array sheet. Following a brief exchange with the interviewer, she redrew it as 10 and 20, before mimicking the arrangement for each box in her box model. While checking to see if her visuals matched her numbers in the written algorithm, Amanda would put one finger from each hand to point to each number on the exterior (either of the array or the box), as her eye-gaze glanced down to the box model for reference (see Figure 2). In each instance, it appears that Amanda focused on the multiplicands but never the relationship between them to produce the product. Later in the interview, when tasked with solving 13×27, Amanda did something similar. However, the interviewer pressed her to explain what 3 times 20 meant with the array, asking "how do you know there are 60 dots?" Like prior tasks, Amanda's eye-gaze focused on the multiplicands but never the interior of the rectangular array that *could* be inferred from her work. She then stated that "3 plus 20 is 23, but if I did it by times, I know that 3 times 2 is 6. Then it's 60 because you add a zero." When asked how she knew there were 21 dots for 3×7, Amanda said "I just know my numbers…my third-grade teacher had these songs that would help us remember…" This interaction confirmed that despite fluency with rotely memorized multiplication facts, Amanda did not understand how to represent or explain multiplication with either single or multi-digit whole numbers.

Figure 2

Amanda's Written Work and Eye-tracking for 15×24

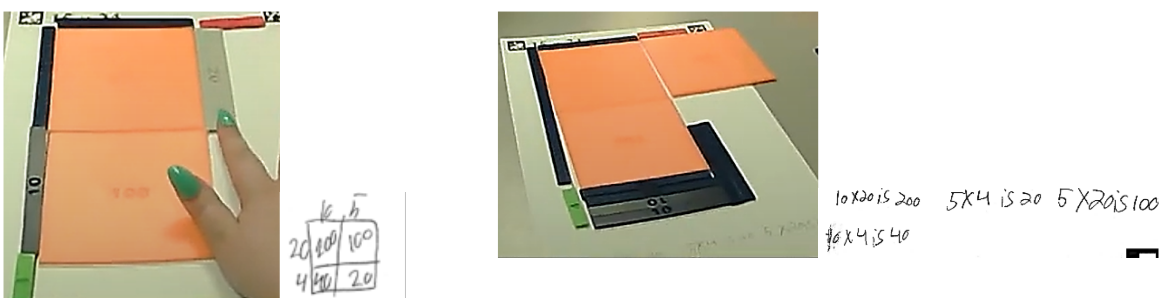


Attend to No More than Two Partial Products

A quarter of participants demonstrated an ability to explain or visually represent their reasoning with no more than two partial products. For example, Kelsey was able to easily model problems like 4×19 and 8×33 with the provided array rods and then explain how the visual represented the problem. When asked to do so with 15×24, Kelsey carefully placed a 10 and 5 rod for 15 at the top, and two 10s and a 4 rod on the left for 24. She then verbally talked through multiplying 10 times 20, looked for and found two 100s before stating “okay and then 5 times 4 is 20” (see Figure 3). After confirming she had finished and was satisfied with her response, the interviewer asked, “have you done this in a different way – a written way?” Kelsey said she had used a box to do so and then was provided with a pencil to show her written approach (see Figure 3). After finishing the written algorithm, Kelsey proceeded to immediately add a 100 block and a 40 block to her visual, but the arrangement did not create a rectangle that coincided with her model. This contrasted her approach when working with tasks requiring only two partial products. Other interview participants demonstrating this theme similarly skipped partial products to multiply, or did not separate a multiplier into factors at all (i.e., 10×24 + 5×24). Thus, regardless of visual used, Kelsey’s approach is representative of actions demonstrated with this theme.

Figure 3

Kelsey (left) and Justin’s (right) Visualizations and Approach to Solving 15×24



Attend to More Than Two Partial Products

A quarter of our participants demonstrated the ability to engage with more than two partial products, indicating a higher level of multiplicative reasoning. Justin was able to visually represent his reasoning on simpler tasks (i.e., 4×19 & 8×33). When prompted to solve 15×24, Justin initially placed two 10s and a 4 on the top and a ten and five on the left but later adjusted so that it would fit on the sheet of paper (see Figure 3). Next, Justin laid out two 100 blocks and began to use 10s to fill in the remaining space. Throughout, Justin’s eye-gaze scanned back-and-forth between the perimeter and area of the array, suggesting a possible coordination between the rods representing area and those representing its dimensions. Halfway through his construction of the area, Justin paused and tapped his pencil while his eye-gaze indicated looking off in the distance. Justin removed five 10s he had placed at the top and replaced them with a 100 (See

Figure 3). We infer from this that Justin anticipated iterating the other five 10s (i.e., 50) and unitized the ten 10s as a 100. Despite this not canonically representing an area model (i.e., the 100 does not line up perfectly with its multiplicands), this was considered as an act of anticipating three levels of units (skip counting the 10s, grouping five 10s as 50, and regrouping/anticipating that two sets of five 10s is one 100s). When prompted to explain his model, Justin wrote "10×20 is 200; 5×4 is 20; 5×20 is 100; 10×4 is 40" and was able to coordinate his written mathematics with the visual representation.

Quantitative Analysis & Results

Following quantitative analysis, the three themes that emerged were quantititized into an ordinal variable (0 = students did not attend to partial products; 1= students attended to no more than two partial products; 2 =students attended to more than two partial products.). Recall that some participants displayed different levels of partial product use in different tasks, so we created composite scores (averages) for these individuals. Quantitization allowed us to examine the correlation between students' MRA scores and their use of partial products, as observed in the interviews. We used the Spearman rank coefficient given our sample size (n = 16). Additionally, the Spearman rho coefficient is ideal for examining ordinal and continuous data (Siegel & Castellan, 1988). Results indicate a statistically significant and strong correlation (ρ = .82, p < .001). Next, we visually examined a scatterplot of the correlated data and observed that students demonstrating MC2 or MC3, as assessed with the MRA, appeared to be those who were consistently attending to partial products during interviews.

Discussion

The purpose of this exploratory study was to examine how children's multiplicative reasoning is related to solving multi-digit multiplication tasks. Results and findings indicated that children operating at the first multiplicative concept or with pre-multiplicative reasoning could not meaningfully use, visually represent, or explain partial products. A primary implication is that students need to be able to anticipate two levels of units or more (MC2 & MC3) to meaningfully understand partial products. These findings highlight a need for instructional approaches to foster students' ability to anticipate and visualize partial products. For example, encouraging both visual and concrete representations is essential for students to gain a conceptual understanding of partial products. Additionally, teachers can provide scaffolding to support students' transitioning from working with two partial products to four partial products.

This study also contributes to understanding how eye tracking technology can provide insights into children's cognitive processes engaged with mathematical tasks. Further study is needed, but this study fills a gap identified by many who have examined multi-digit multiplication (Izsàk, 2005; Hickendorff et al., 2019).

Acknowledgments

The research reported here received support from the National Science Foundation (NSF) through DRK-12 Grant #2404553. Any opinions, findings, conclusions, or recommendations in this paper are those of the author(s) and do not necessarily reflect the views of NSF.

References

Duchowski, A. (2007). Eye tracking methodology. Theory and practice. Springer.

Fuson, K. C. (2003). Toward computational fluency in multidigit multiplication and division. *Teaching Children Mathematics*, *9*(6), 300-305.

Hackenberg, A. J. (2010). Students' reasoning with reversible multiplicative relationships. *Cognition and Instruction, 28*(4), 383-432.

Hickendorff, M., Torbeyns, J., & Verschaffel, L. (2019). Multi-digit addition, subtraction, multiplication, and division strategies. *International handbook of mathematical learning difficulties: From the laboratory to the classroom*, 543-560.

Hurst, C., & Hurrell, D. (2016). Multiplicative thinking: much more than knowing multiplication facts and procedures. *Australian Primary Mathematics Classroom, 21*(1), 34-38.

Iszák, A. (2005). You have to count the squares: Applying knowledge in pieces to learning rectangular area. *Journal of Learning Sciences, 14*(3), 361–403.

Kosko, K. W. (2019). A multiplicative reasoning assessment for fourth and fifth grade students. *Studies in Educational Evaluation*, *60*, 32–42.

Larsson, K., Pettersson, K., & Andrews, P. (2017). Students' conceptualisations of multiplication as repeated addition or equal groups in relation to multi-digit and decimal numbers. *The Journal of Mathematical Behavior*, *48*.

Lin, F. Y., & Kubina, R. M. (2005). A preliminary investigation of the relationship between fluency and application for multiplication. *Journal of Behavioral Education, 14*, 73-87.

Norton, A., Boyce, S., Phillips, N., Anwyll, T., Ulrich, C., & Wilkins, J. (2015). A written instrument for assessing students' units coordination structures. *Journal on Mathematics Education*, *10*(2), 111–136.

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye Movements as Reflections of Comprehension Processes in Reading. *Scientific Studies of Reading, 10*(3), 241–255.

Young-Loveridge, J., & Mills, J. (2009). Teaching multi-digit multiplication using array-based materials. In R. Hunter, B. Bicknell, & T. Burgess (Eds.), Crossing divides: *Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia* (Vol. 2). MERGA.

ACCURACY, EFFICIENCY, AND FLEXIBILITY ACROSS MULTIPLICATIVE DOUBLE COUNTING STAGES

Bridget Broome
Oklahoma State University
bridget.broome@okstate.edu

This paper examines the relationship between elementary students' multiplicative double counting (mDC) stage and each concept of fluency. Students' stage of mDC was identified as either pre-multiplicative, participatory, or anticipatory. In addition, a multiplication fluency assessment that measured accuracy, efficiency, and flexibility was given. Results indicated pre-multiplicative students' overall score on each concept of fluency were lower than the other stages and anticipatory students demonstrated the highest levels of efficiency and flexibility. However, their scores on these two concepts were lower than anticipated, so future research should focus on ways to increase efficiency and flexibility for this population of students.

Fluency is defined by the National Council for Teachers of Mathematics (NCTM) (2014) as the ability to use procedures accurately, efficiently, and flexibly and to apply the procedures to new problems and contexts. Kling and Bay-Williams (2015) apply this definition of fluency to multiplication facts and state that developing fluency with multiplication facts is one of the most important objectives of teachers in grades three through five. However, they recognize that mastery of these facts continues to be a major stumbling block for many elementary students. Fluency with multiplication facts impacts students' comprehension of multiple mathematics concepts including fractions. Therefore, one goal of teaching multiplication to students should be for them to become fluent, and in order for students to obtain fluency, as defined by the NCTM, all three concepts of fluency (i.e., accuracy, efficiency, and flexibility) must be developed.

Another area where students continue to struggle in elementary grades is with their ability to reason multiplicatively. Tzur et al. (2013) propose a developmental framework for multiplicative schemes that students construct as they develop multiplicative reasoning. The first scheme of the developmental framework is multiplicative double counting (mDC) which "involves recognizing a given number of composite units, each consisting of the same number of 1s" (p. 90). They explain schemes as "conceptual structures and operations children construct and use for reasoning in multiplicative situations" (p. 85). Once students have established an mDC, they can use strategies for unknown multiplication facts. For example, a student who has constructed an mDC may solve 8×6 by using the known fact of 8×5=40 and adding one more group of 8 to get

48. In the same way a student who is fluent, as defined by NCTM (2014), would also be able to solve this problem in a similar way by demonstrating accuracy, efficiency, and flexibility.

Literature Review

Bay-Williams and SanGiovanni (2021) state that when students can complete problems correctly, they are often deemed fluent. They argue, however, that a fluent student would not just answer correctly, but can explain how they reasoned and what strategies they could use to solve the problem. They propose that fluency is much more involved than just being able to answer a math problem accurately and should include efficiency and flexibility.

Of the three areas of fluency, accuracy is the most easily defined because it is straightforward. *Accuracy* is "correctly solving a procedure" (Bay-Williams & SanGiovanni, 2021, p. 3). If a student gets an answer correct, then they will be considered accurate. Because it is the most straightforward, it is also the easiest to assess and therefore, prior research has focused more on accuracy than efficiency and flexibility. Bay-Williams and SanGiovanni (2021) define *efficiency* as a student's ability to select a strategy that is appropriate for the problem. These students know which strategy to use and implement quickly. To be considered efficient, students must work through the problem without getting stuck. The final concept of fluency is flexibility, and it is the most challenging to define and therefore, assess. According to Bay-Williams and SanGiovanni (2021), a student who demonstrates *flexibility* knows several strategies for solving a problem and can apply these strategies to new problem types. They are able to switch approaches when needed and can generalize strategies for various situations. Prior research on fluency has only focused on speed and accuracy. This study seeks to view fluency through the theoretical lens of mDC to better understand the struggles that students are facing as they learn their multiplication facts.

Theoretical Framework

There are three stages of mDC. The first of these stages is the pre-multiplicative stage. Children who are at this stage typically use additive reasoning when solving multiplicative situations and count using something concrete to keep track of additional counts (Tzur et al., 2021). Students are considered to be at the pre-multiplicative stage when they cannot figure out the answer to a multiplicative situation in any way other than using a counting by ones strategy. Students that use counting by ones strategy often find the correct answer to a multiplication problem, however, they are inefficient and do not demonstrate flexibility. The next stage of mDC is the participatory stage. Students at the participatory stage are in the process of transitioning from additive to multiplicative reasoning (Steffe, 2010). Because of this they tend to revert to additive reasoning when solving multiplication tasks. Ulrich (2015) states that students reasoning at this stage experience significant constraints in dealing with the added complexity of coordinating two-levels of units and this causes them to revert back to additive reasoning by using strategies such as skip counting or repeated addition. Although students often get an accurate answer when using strategies such as skip counting and repeated addition, these strategies are not always the most efficient or flexible. The most sophisticated stage of mDC is that of the anticipatory mDC. Students at this stage see tasks as multiplicative and not additive and can reason strategically about multiplication (Tzur et al., 2021). This allows them to use efficient and flexible strategies such as decomposition and compensation when solving multiplication problems. Therefore, this study seeks to answer the question: How does a student's mDC stage relate to their accuracy, efficiency, and flexibility with multiplication facts?

Methods

Participants and Sampling Strategy

Approval of this study was given by the IRB at Oklahoma State University (IRB-23-561). Participants for this study were selected from a rural midwestern elementary school. All students in fourth and fifth grade were eligible to participate (n=65). Of the students who took the assessments there were 28 fourth graders and 31 fifth graders that took both assessments and had permission and gave assent to have their assessments scored for use in the study. The final sample size was 59 students. Each student that participated was told to answer each question as best they could and to not erase any of their work or skip any questions. They wrote "I don't know" if they could not solve a problem. The assessments were not timed.

Quantitative Instruments

Students were asked to complete two quantitative assessments as part of the data collection. One assessment measured their mDC stage and one measured their multiplication fluency level. The mDC assessment measured students' mDC stage as either pre-multiplicative, participatory, or anticipatory. This assessment was previously validated (Tzur et al., 2022) and was scored based on correct and incorrect answers. Raw data was entered into spreadsheets that was provided through personal communication (R. Tzur, personal communication, February 10, 2023). After entering the scores, each student was assigned an mDC stage (0- pre-multiplicative, 1- participatory, and 2- anticipatory).

An additional instrument was used to assess students' multiplication fluency level. Because an instrument was not available that measured all three concepts of fluency, a written fluency assessment was created specifically for this study and content validation was obtained before its use. A pilot study was conducted and included a sample of convenience consisting of six participants from third and fourth grade (3rd grade, n = 2; 4th grade, n=4). This iteration of the multiplication assessment included six multiplication problems. The problems included foundational facts (i.e., 6 × 2) as well as math facts that would elicit student strategy use such as 4 × 9. Students could calculate 4 × 9 either through doubling 2 × 9 or through compensation using (5 × 9)−9. The assessment also included one 2-digit by 1-digit multiplication problem (22 × 5) to evaluate whether students would use the standard algorithm or partial products to find the product. Students were asked how they solved the problem, how they could explain to a friend how to solve the problem, and what strategies they used or could use to solve the problem. These questions were used so that students' use of strategy could be seen and also to see whether they could solve the problems using more than one strategy.

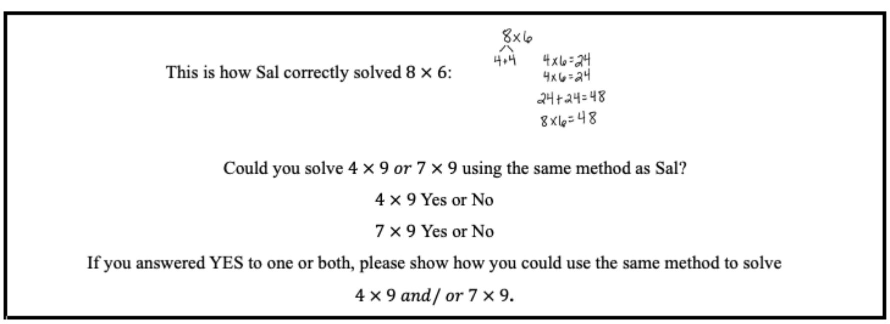
While scoring the first iteration of the assessment, it was determined that there were no questions that allowed researchers to assess flexibility with multiplication. Because of this, five questions that included worked examples were added. The worked examples demonstrated compensation, doubling, near squares and partial products. Another edit made included a part A and part B for each of the first five problems with each part being on a separate page. In part A students solved the given multiplication problem and explained how they solved it. Part B showed students one possible strategy they could use and asked the question, "Is this the BEST way to solve the problem? Why or why not?" followed by "Is there another way to solve the problem?" Parts A and B were on separate pages to ensure that students did not go back to the previous section and solve the problem in the way that was shown to them. Two experts in the field of multiplicative reasoning and fluency, Karl Kosko[1] and John SanGiovanni[2] provided feedback on this second iteration of the multiplication fluency assessment. Kosko suggested including an application of the associative property as one option for a question. Taking this feedback into consideration, the worked example of 8 × 7 was replaced with a worked example of 4 × 7 to elicit students' use of the associative property by looking for them to reconsider solving 4

× 7 by calculating 2 × 14 (i.e., halving and doubling). He indicated that he believed the items on the assessment would yield interesting data about students' fluency and understanding of multiplication (K. Kosko, personal communication, January 25, 2024). SanGiovanni indicated that the assessment would measure student's multiplication fluency. His only critique was that beginning with a problem such as 6 × 2 might "muddy the water" for students when asking if they know an alternative strategy for solving the problem since it is a foundational fact that most fourth and fifth graders have memorized (J. SanGiovanni, personal communication, January 16, 2024). Taking this feedback into consideration, 6 × 2 was made a sample problem on the assessment and it was only scored for accuracy and efficiency but not flexibility.

The final version of the assessment included 10 multiplication problems that allowed students to show their understanding of multiplication and explain how they solved multiplication fact problems using various strategies. Each question on the fluency assessment was scored individually with a score of zero or one on each concept of fluency. Students were given an accuracy score of one if they had the correct answer to the multiplication problem regardless of how they obtained their answer. Because of the way the fluency assessment was created, there were 13 opportunities for students to demonstrate accuracy. There were three questions that listed two multiplication problems. These questions asked which multiplication fact problem could best be solved using the method shown in the worked example (Figure 1).

Figure 1

Fluency Assessment Problem 5


This is how Sal correctly solved 8 × 6:

8×6
4+4
4×6=24
4×6=24
24+24=48
8×6=48

Could you solve 4 × 9 *or* 7 × 9 using the same method as Sal?

4 × 9 Yes or No

7 × 9 Yes or No

If you answered YES to one or both, please show how you could use the same method to solve

4 × 9 *and/ or* 7 × 9.

It was not necessary that students solve both problems, however, some did. To account for this, attempts on each problem were tabulated to determine how many problems out of 13 each student attempted. Students were not penalized for problems where there was no attempt. The total correct was divided by the total attempted in order to give students an overall percentage for accuracy. Scoring efficiency and flexibility was more straightforward as there were 10 opportunities and students were given a score of zero or one on each of the 10. A score of one indicated that they demonstrated fluency skill and a score of zero indicated they did not demonstrate fluency skill. To evaluate efficiency, the rubric helped determine whether the student chose a strategy suitable for the given problem. A scoring rubric was created based upon anticipated student answers. An example of the scoring rubric is shown in Table 1. The results of the fluency assessment were broken down into categories for overall accuracy, efficiency, and flexibility percentages in order to see which concept was most associated with mDC stage.

Table 1

Example of Scoring Rubric

Scoring Rubric for Multiplication Fact Fluency Assessment – Problem 4 x 9 or 7 x 9			
Note for Accuracy: Accuracy is scored for each problem on these items. If both problems are attempted, a score of 1 will be given on attempted and if they are both correct a score of 1 will be given for both problems. If only one is attempted, then a score of 1 is only given for the problem attempted and a score of 0 will be given for the problem that was not attempted.			
Efficiency		**Flexibility**	
1	0	1	0
-Yes to both, uses any strategy correctly. -Yes to 4 x 9 (correctly) and No to 7 x 9 (no attempt) -Yes to 7 x 9 (correctly) and no to 4 x 9 (no attempt)	-No for 4 x 9 and 7 x 9 -Yes for both but can't use correctly	-Yes to both, uses given strategy correctly -Yes to 4x9 (uses strategy given) and No to 7x9 (no attempt)	-No for 4x9 -Yes for both but can't use given strategy correctly

Results

The results of the mDC assessment and accuracy, efficiency and flexibility scores are shown in Table 2. From the table it is evident that pre-multiplicative mDC students and participatory mDC students' scores on efficiency and flexibility are lower than those of the anticipatory mDC students. While the anticipatory mDC students scored higher in every concept, their scores on efficiency and flexibility were higher. Kendall's Tau b was calculated in order to determine which concepts of fluency were most associated with mDC stage. The results of the statistical analysis indicated a significant relationship between mDC stage and each concept of fluency. The results for significance between accuracy and mDC stage indicate a weak association (τ_b=.286, p=.026). Efficiency (τ_b=.373, p<.001), and flexibility (τ_b=.336, p<.001) were moderately associated with mDC. Robust confidence intervals also indicate a significant relationship between each concept and mDC stage.

Table 2

Overall Accuracy, Efficiency, and Flexibility Percentages

			Overall Accuracy Percentage			
mDC Stage	0-20%	21-40%	41-60%	61-80%	81-100%	Total
Pre-multiplicative	0%	7.1% (1)	21.4% (3)	28.6% (4)	43% (6)	100% (14)
Participatory	0%	0%	0%	28.6% (4)	71.4% (10)	100% (14)
Anticipatory	0%	0%	9.7% (3)	6.5% (2)	84% (26)	100% (31)

			Overall Efficiency Percentage			
mDC Stage	0-20%	21-40%	41-60%	61-80%	81-100%	Total
Pre-multiplicative	50% (7)	21.4% (3)	28.6% (4)	0%	0%	100% (14)
Participatory	28.6% (4)	21.4% (3)	28.6% (4)	14.3% (2)	7.1% (1)	100% (14)
Anticipatory	9.7% (3)	22.6% (7)	42% (8)	29% (9)	12.9% (4)	100% (31)

			Overall Flexibility Percentage			
mDC Stage	0-20%	21-40%	41-60%	61-80%	81-100%	Total
Pre-multiplicative	71.4% (10)	28.6% (4)	0%	0%	0%	100% (14)
Participatory	64.3% (9)	21.4% (3)	14.3% (2)	0%	0%	100% (14)
Anticipatory	35.5% (11)	16.1% (5)	25.8% (8)	22.6% (7)	0%	100% (31)

Note. Number of students in parentheses.

Discussion

These findings demonstrate that efficiency and flexibility are more strongly associated with a student's mDC stage than accuracy. Although accuracy did increase as mDC stage increased, the increase in efficiency and flexibility is more evident. These results demonstrate that at least 81% accuracy is associated with students at any stage of mDC, however, an overall percentage of 60% or higher on efficiency and flexibility was only associated with students at a more sophisticated stage of mDC. Anticipatory mDC students demonstrated the highest overall percentages for each concept of fluency. These students are able to reason multiplicatively and operate on composite units allowing them to use strategies that are both efficient and flexible. This is a novel contribution of this study and adds to the literature on multiplicative reasoning and fluency level as this relationship has not appeared in prior research. This is an important contribution because if fourth and fifth graders are unable to reason multiplicatively, their ability to demonstrate fluency could be greatly impacted. This not only contributes to their ability to be fluent with their multiplication facts, but students who are not fluent with the multiplication facts also struggle with fractional concepts which are prevalent across fourth and fifth grades.

Conclusion

Although the results of this research show that mDC stage is related to each concept of fluency with this association being stronger between mDC stage and efficiency and flexibility, there is still work to be done. According to these results, anticipatory mDC students did not score above 80% overall on flexibility and only 12.9% scored between 81-100% overall on efficiency. These students are ready to receive instruction that allows them to demonstrate efficiency and flexibility, however, many of them are not demonstrating their ability to reason in this way. Future research should focus on what can be done to increase these two concepts in students that are at the participatory mDC stage or higher.

References

Bay-Williams, J. M., & SanGiovanni, J. (2021). *Figuring out fluency in mathematics teaching and learning, grades K-8: Moving beyond basic facts and memorization*. Corwin.

Kling, G., & Bay-Williams, J. M. (2015). Three steps to mastering multiplication facts. *Teaching Children Mathematics*, *21*(9), 548–559. https://doi.org/10.5951/teacchilmath.21.9.0548

National Council of Teachers of Mathematics (NCTM). (2014). *Principles to actions: Ensuring mathematical success for all*. NCTM.

Steffe, L. P. (2010). Operations that produce numerical counting schemes. In L. P. Steffe & J. Olive (Eds.), *Children's Fractional Knowledge* (pp. 27–47). Springer.

Tzur, R., Johnson, H. L., McClintock, E., Kenney, R. H., Xin, Y. P., Si, L., Woodward, J., Hord, C., & Jin, X. (2013). Distinguishing schemes and tasks in children's development of multiplicative reasoning. *PNA. Revista de Investigación En Didáctica de La Matemática*, *7*(3), 85–101. https://doi.org/10.30827/pna.v7i3.6128

Tzur, R., Johnson, H. L., Norton, A., Davis, A., Wang, X., Ferrara, M., Harrington, C., & Hodkowski, N. M. (2021). Children's spontaneous additive strategy relates to multiplicative reasoning. *Cognition and Instruction*, *39*(4), 451–476. https://doi.org/10.1080/07370008.2021.1896521

Tzur, R., Johnson, H. L., Davis, A., Hodkowski, N. M., Jorgensen, C., Wei, B., & Norton, A. (2022). A stage-sensitive written measure of multiplicative double counting for grades 3-8. *Studies in Educational Evaluation*, *74*, 1–14. https://doi.org/10.1016/j.stueduc.2022.101152

Ulrich, C. (2015). Stages in constructing and coordinating units additively and multiplicatively (part 1). *For The Learning of Mathematics*, *35*(3), 2–7. https://doi.org/4438267

WHEN MATH APPS MISS THE MARK: THE IMPACT ON A STUDENT'S MATH IDENTITY AND MOTIVATION

Micah Swartz
Texas State University
mas762@txstate.edu

Math apps are increasingly prevalent in classrooms, yet their impact on students' math identity and motivation is not well understood. This case study examined Sarah, a third grader, and how her interactions with math apps influenced her math identity and motivation. Findings revealed that Sarah's enjoyment of math apps was contingent on the alignment between her perception of math and the apps' emphasis on computational approaches. This study contributes to ongoing discussions about the intersection of technology and student disposition and underscores the importance of considering individual student experiences in technology integration.

"Being on laptops is not good for us." This quote was from a recent interview with a third grader, Sarah, who shared her thoughts on using math apps. Over the last few decades, math applications (e.g., IXL, Prodigy, ST Math, Zearn, etc.) have become increasingly prevalent in K-12 schools worldwide and have been acknowledged as an influential part of learning mathematics (Griffith et al., 2020; Laato et al., 2020). Recent research on math apps has largely focused on the design and content of apps (Laato et al., 2020) or research on academic outcomes and achievement-related metrics (Griffith et al., 2020).

Since elementary students are a common target group for mathematics apps (Laato et al., 2020), there is a high need to study this population of students. Existing literature indicates that identity is an important aspect of students' learning experiences as it impacts their success and well-being in the classroom (Bohrnstedt et al., 2021). Similarly, motivation is a well-established area of research in mathematics education, and motivation has been recognized as an important mediator of mathematics learning and achievement (Schukajlow et al., 2023). This presentation examines the case of Sarah, a third-grader, and explores the following questions: *What is the relationship between math apps and Sarah's math identity and motivation? How did Sarah's math identity and motivation change over time spent using math apps?*

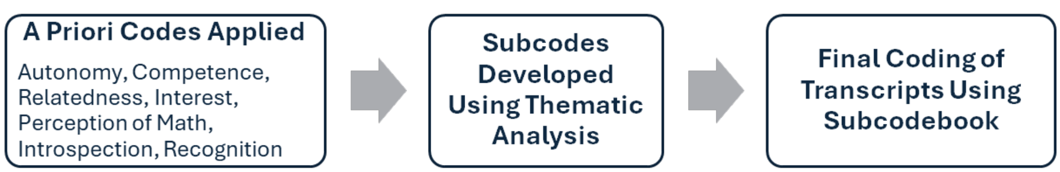
Methods

The case of Sarah is part of a larger exploratory multiple-case study (Yin, 2016) that examined third graders' math identity and motivation related to their weekly use of math apps. This multiple-case study involved eight cases, each carefully selected from a classroom participant pool to represent a variety of mathematical identities and motivational profiles. For this preliminary analysis, the focus was placed on Sarah's case, as it offered particularly insightful reflections and was especially revealing (Yin, 2016). Previous work has focused on students' math identity and motivation in separation, looking at how each construct independently relates to math app use (Swartz, 2024a, 2024b), but this paper examines Sarah holistically, analyzing her math identity and motivation in conjunction.

To operationalize math identity, I draw on the mathematics identity framework proposed by Cribbs et al. (2015), which views math identity as comprising of three constructs: Perception of Math (originally Interest but renamed to avoid confusion with the motivation construct Interest), Introspection (originally Competence/Performance but renamed to simplify things), and Recognition. Perception of Math denotes a student's view of math, Introspection describes a student's view of themselves related to math, and Recognition denotes a student's perception of how others recognize them as a math student. To operationalize motivation, I utilize self-determination theory (SDT), which asserts that individuals share three fundamental psychological needs: autonomy, competence, and social relatedness (Ryan & Deci, 2000). Autonomy involves having control over one's learning, competence refers to feeling capable of achieving desired outcomes, and social relatedness means feeling connected to others. My application of SDT leverages the unique opportunities math apps provide to foster these needs. I also consider the interest task value from expectancy-value theory (Eccles et al., 1983) since my research shows that students enjoy math games and apps (Shin et al., 2012).

Figure 1

Coding Process



Data gathered for Sarah's case includes two interviews, weekly surveys, observations, and math app data. I applied a priori coding (Miles et al., 2014) using categories like Perception of Math, Introspection, and Recognition for math identity, and Autonomy, Competence, Relatedness, and Interest for motivation. Using thematic analysis (Braun & Clarke, 2012), I created subcodes to explain each case's math identity and motivation. After developing a subcodebook (examples in Tables 1 & 2), I assessed coding reliability by calculating the agreement between myself and two independent experts, in math identity and motivation. A subset of 30% of the dataset was randomly chosen; following detailed discussions on the sub codebook and training on 15% of the data, both secondary coders and I independently coded 15% of the data, achieving interrater agreements of 99% for both codebooks.

Findings

"Sarah is an old soul. She is one of the sweetest kids and is a hard worker. She is extremely bright." – Ms. Care, Sarah's Teacher. Sarah was an introverted third grader who preferred quietly reading a book in the corner over group work, that is, if she was not paired up with Madison or Chloe, her two best friends. Sarah could be especially withdrawn and inward-focused since her peers were often boisterous during moments of free time and group work. While Sarah was often quick to retreat to a reserved, independent activity, she was also not afraid to voice her opinion when she felt confident or was encouraged by her friends. She had recently transferred to Cedar Hill Elementary School, where over 60% of the school's students scored at or above the proficient level on the state math test, nearly double the state average. Sarah described herself as female and Caucasian, and institutionally, her school labeled her as being gifted and talented. To examine the relationship between Sarah's math identity and motivation and her use of math apps and understand how her math identity and motivation changed over the duration of the study, I focus on the components of Sarah's math identity and motivation that experienced major shifts. In the following analysis, I separated two types of math Sarah engaged in: regular math (math free from technology) and math app math (the math in math apps).

Changes in Identity

While some aspects of Sarah's math identity changed minimally or not all, other components changed in more obvious ways (highlighted in grey). Table 1 summarizes the components of Sarah's math identity based on her interviews in January and May of 2024. Examining Sarah's math identity at the beginning and end of the semester, we see no changes in how she perceived others to view her math ability (Recognition) and how she viewed her own math ability (Introspection). However, we do see changes in Sarah's view of regular math and math apps (Perception of Math—see text highlighted in gray). Since only an observed change in the component of Sarah's math identity related to her view of math (Perception of Math) was observed, I now further explore the changes seen in the component across the semester.

Perception of Math

In January, Sarah viewed math as doing facts and getting to the answer. In displaying a more computational view of math, Sarah felt that part of what it means to do math is to get the answer quickly. Responding to the Screening Tool question, “What does it mean to be good at math?” Sarah said, “To not only get the answers right, but to also answer them quickly.” There was an element of speed engrained in how she viewed doing math. In the first interview, she reiterated this feeling when she stated, “That’s part of the point of math, to just get them [facts] right.” This belief was likely influenced by her view that math app math was fast, often rushing her to complete problems and facts before she could fully think through the problem. Sarah said, “Reflex, you have to get it really fast, and you have to get it right.” Some math apps emphasized solving problems quickly, which seemed to translate to her viewing regular math as solution-oriented, where the goal of math is to find answer and solve the fact.

Table 1

Overview of Changes in Sarah’s Math Identity

		Interview 1	Interview 2
Regular Math Identity Components & Most Prevalent Themes (in descending order of prevalence)	**Perception of Math**	(1) Math is exciting/fun.	(1) Math is exciting/fun.
		(2) Math is doing facts/getting the answer.	(2) Math is solving and making sense of problems.
			(3) Math is doing/knowing facts.
	Introspection	(1) Good at math.	(1) Good at math.
	Recognition	(1) Good at math.	(1) Good at math.
Math App Math Identity Components & Most Prevalent Themes (in descending order of prevalence)	**Perception of Math App Math**	(1) Math app math is exciting/fun.	(1) Math app math is not exciting/fun.
		(2) Math app math is a game/not real math.	(2) Math app math is fast.
	Introspection	(1) Good at math	(1) Good at math
	Recognition	(1) Good at math.	(1) Good at math.

By the end of the semester in May, Sarah displayed a more problem-solving view of math and felt that part of what it meant to do math is to understand the solution you’ve found. In responding to the Screening Tool question, “What does it mean to be good at math?” Sarah said, “To get the answers right and understand them.” When asked to define what math was in the second interview, Sarah replied, “It’s a way of understanding numbers and solving problems to help you understand, like how many stars are on the American flag or something.” In May, Sarah’s view of math emphasized understanding and problem-solving, marking a stark shift from her view of math in January as computational and driven by answering facts quickly.

The second major shift in Sarah’s perception of math came in her view of math app math. In January, Sarah primarily viewed math app math as exciting/fun and a gamified experience that she did not consider to be "real math." By May, Sarah found math app math unexciting and felt it focused on solving facts quickly. Perhaps Sarah no longer had fun doing math app math because of the shift in her perception of what math is. When her perception of math and the math emphasized in math apps were congruent (i.e., getting facts right), Sarah experienced enjoyment when doing math app math. However, when her perception of math shifted away from a perspective that stressed a computational and rote view toward a view foregrounding understanding and making sense of solutions, Sarah no longer experienced enjoyment as the dominant feeling when engaging in math app math. In fact, it was quite the opposite, with feelings of "frustration" and being "grumpy" due to the focus on efficiency through the app's timer feature and a focus on correctness through the app's rating feature, which tracked students' fact fluency percentage and as Sarah described, "It just tells you how good you are."

Changes in Motivation

Examining changes in Sarah's motivation (see Table 2), the most evident changes in her motivation are seen in a value increase in relatedness from low to high and a complete reversal of the enjoyment comparison between regular math and math app math. More minimal changes were observed with an increase in the value of autonomy, a decrease in the value of competence, and changes in the experience of both autonomy and relatedness.

Relatedness

Relatedness was a motivation component that Sarah valued more at the end of the semester than at the beginning. At the beginning, Sarah felt it was "not really" that important to feel close and connected with her peers when doing math. By the end of the semester, Sarah valued feeling connected to classmates and seemed to differentiate opportunities for connection in regular math and math app math based on the quality of interaction. While math apps like Prodigy and Boddle allowed students to experience relatedness by battling other students online, the quality of interaction via math apps had a low value for Sarah, and she expressed a high value and enjoyment in interacting face-to-face with her peers. At the end of the semester, Sarah ranked group work without technology as her favorite way to do math.

Interest

Perhaps the most evident shift in Sarah's motivation came when she reversed the order of which type of math she enjoyed (Interest) more (Table 2). While the interest Sarah felt when doing regular math was consistently high throughout the semester, Sarah started the semester holding the view that math app math was more enjoyable than regular math. For Sarah, many things about math apps were enjoyable: the ability to customize characters' outfits, the ability to buy items for the virtual room her game character resided in, and the animations and sounds the game made as she solved problems. At one point during the first interview, she pulled off her headphones while playing Reflex and said, "This sounds really funny." Sarah acknowledged that while the enjoyment she experienced with math apps was greater than that of regular math, this enjoyment was problematic because there wasn't enough "work," which I interpret as Sarah problematizing the gamified nature of some of the apps.

Table 2

Overview of Changes in Sarah's Motivation

		Interview 1	Interview 2
Value on Motivation Component	**Autonomy**	Moderate	Moderate
	Competence	High	Moderate
	Relatedness	Low	High
Experience (Regular Math vs Math app Math)	**Autonomy**	Math apps and regular math offer similar autonomy.	Regular math offers greater autonomy.
	Competence	Regular math and math apps offer similar competence.	Regular math and math apps offer similar competence.
	Relatedness	Regular math and math apps offer similar relatedness.	Regular math offers greater relatedness.
	Interest	Math apps are more fun.	Regular math is more fun.

At the end of the semester of regularly utilizing her laptop to engage in math apps, Sarah said, "I love all no computer [days]" since she felt that math apps are "not that much fun." Sarah explained that she felt a major reason was that her classmates often had behavior issues when on their computers playing math apps. Describing several reasons she no longer liked math apps, she said, "I mean, sometimes it's just not that much fun. It makes people scream a lot because you can interact so much, and it's a lot more games than math. And some of the math is really, really easy." Another aspect of math apps that Sarah became turned off by was the timing mechanism on many learning apps. For example, Reflex had a timer counting down as students attempted math facts, or students would be challenged to complete as many facts in a set time as possible. Describing how she felt when solving problems as these timers counted down, she said, "stressed and frustrated." She experienced an apparent decline in enjoyment related to math apps over the duration of the study.

Discussion

Sarah's evolving math identity and motivation offers a microcosm of how math apps relate to students' math identity and motivation. For instance, Sarah's enjoyment (Interest) of math apps appeared tied to her evolving Perception of Math. When her Perception of Math aligned with the computational emphasis in these apps, she experienced more enjoyment when engaging with math apps. However, as her Perception of Math shifted toward a focus on understanding and problem-solving, her enjoyment of math apps diminished. This highlights a significant issue with many current math apps: their heavy reliance on gamified features and time-based mechanics often fails to accommodate deeper mathematical thinking and problem-solving.

Sarah's case also offers valuable insights for educators integrating technology into mathematics instruction. One notable finding is the decline in Sarah's enjoyment of math apps and her increased preference for regular math, particularly group work. This shift underscores the importance of social interaction in fostering motivation and engagement. While apps like Prodigy and Boddle allow for virtual interaction, Sarah's comments suggest that face-to-face collaboration offers a deeper sense of relatedness. Educators should be mindful of this when incorporating math apps into their classrooms and consider pairing app-based activities with opportunities for peer interaction and collaborative problem-solving.

Sarah's experiences point to areas where math apps could be improved to better support students' math identity and motivation. Learning app developers should critically evaluate the role of gamification. While features like customizable avatars and rewards can make apps appealing, they should not overshadow the mathematical content. Games that integrate meaningful mathematical challenges with engaging narratives, rather than superficial gamified elements, may do better at sustaining long-term interest while supporting a positive math identity. This study focuses on Sarah's case, which raises questions about the broader impact of math apps on elementary students' math identity and motivation. Future research should determine if the patterns seen in Sarah's case are consistent in a larger, more diverse student sample. Cross-case analyses could reveal whether the features Sarah had a disdain for—like speed emphasis and lack of meaningful interaction—are systemic issues in math app design. This study connects her case to broader trends, contributing to discussions about technology's role in math education. It highlights the need for intentional math app design that engages students and aligns with their evolving identities and motivational needs. Such efforts are vital to ensure technology fosters meaningful learning instead of hindering deeper mathematical understanding.

Acknowledgments

This material is based on work supported in part by the National Science Foundation (NSF) under grant number DGE-2235199. Any opinions, findings, and recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NSF.

References

Bohrnstedt, G. W., Cohen, E. D., Yee, D., & Broer, M. (2021). Mathematics identity and discrepancies between self-and reflected appraisals: their relationships with grade 12 mathematics achievement using new evidence from a US national study. *Social Psychology of Education, 24*(3), 763–788. https://doi.org/10.1007/s11218-021-09631-0

Braun, V., & Clarke, V. (2012). *Thematic analysis*. American Psychological Association.

Cribbs, J. D., Hazari, Z., Sonnert, G., & Sadler, P. M. (2015). Establishing an explanatory model for mathematics identity. *Child development, 86*(4), 1048–1062. https://doi.org/10.1111/cdev.12363

Eccles J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146).

Griffith, S. F., Hagan, M. B., Heymann, P., Heflin, B. H., & Bagner, D. M. (2020). Apps as learning tools: a systematic review. *Pediatrics, 145*(1). https://doi.org/10.1542/peds.2019-1579

Laato, S., Lindberg, R., Laine, T. H., Bui, P., Brezovszky, B., Koivunen, L., ... & Lehtinen, E. (2020). Evaluation of the pedagogical quality of mobile math games in app marketplaces. In *2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)* (pp. 1–8). IEEE. https://doi.org/10.1109/ice/itmc49519.2020.9198621

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook*. Sage.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, *55*(1), 68. https://doi.org/10.1037//0003-066x.55.1.68

Schukajlow, S., Rakoczy, K., & Pekrun, R. (2023). Emotions and motivation in mathematics education: Where we are today and where we need to go. *ZDM*, 1–19. https://doi.org/10.1007/s11858-022-01463-2

Shin, N., Sutherland, L. M., Norris, C. A., & Soloway, E. (2012). Effects of game technology on elementary student learning in mathematics. *British journal of educational technology*, *43*(4), 540–560. https://doi.org/10.1111/j.1467-8535.2011.01197.x

Swartz, M. (2024a). Exploring Motivation and Math Apps: A Third Grader's Story. In D. Kombe, & A. Wheeler (Eds.), *Proceedings of the 51st Annual Meeting of the Research Council on Mathematics Learning*. (pp. 65–73). RCML, Columbia, SC.

Swartz, M. (2024b). Math Identity and Math Apps: What in Common?. In Kosko, K. W., Caniglia, J., Courtney, S. A., Zolfaghari, M., & Morris, G. A. (Eds.), *Proceedings of the forty-sixth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. (pp. 1678–1683). PMENA, Cleveland, OH. https://www.doi.org/10.51272/pmena.46.2024

Yin, R.K. (2016). *Qualitative Research from Start to Finish, Second Edition.* The Guilford Press.

VALIDATION OF A SHORTENED MEASURE OF STUDENTS' BELIEFS ABOUT PROBLEM SOLVING

Sam Rhodes
Georgia Southern University
srhodes@georgiasouthern.edu

Antonio Gutierrez de Blume
Georgia Southern University
agutierrez@georgiasouthern.edu

Rick Bryck
Landmark College
rickbryck@landmark.edu

Patrah Frimpong
Georgia Southern University
pf03994@georgiasouthern.edu

This paper shares shortened versions of three of the Indiana Mathematics Beliefs Scales. Exploratory Factor Analysis (EFA) with principal axis factoring extraction method (common factor extraction) and a promax (oblique) rotation were used to analyze data from a sample of 227 middle school students. The results showed that the scales could be shortened from a total of 18 items to 11 items while retaining adequate internal consistency.

Background

Beliefs may be broadly defined as assumptions or propositions held by individuals that they consider to be true and that may implicitly or explicitly influence actions (Sun & Zhang, 2024; Voss et al., 2013). Research studies that have explored mathematical beliefs often fall into one of five categories of focus: understanding beliefs about the nature and structure of mathematics, understanding how beliefs evolve or change over time, exploring how different beliefs impact behavior and achievement, understanding differences in beliefs across various domains, and testing interventions aimed at changing the beliefs that students have about themselves and the subject (Iannone & Simpson, 2019; Muis, 2004).

Within this research base, numerous studies have found that beliefs may impact students' motivation, comprehension of mathematical texts, and achievement (Iannone & Simpson, 2019; Mcleod, 1992; Schommer et al., 1992; Schunk, 1991). In addition, research has specifically explored the role of beliefs when solving cognitively demanding tasks and problems. For example, Chapman (2015) summarized research across seminal studies on problem solving (PS) - defined herein as the pursuit of a mathematical goal when the solution is not immediately clear (Lester, 2013)—to suggest that beliefs and dispositions played an important role in PS success. This claim was later supported by Rhodes et al. (2023) who conducted a multiple regression of factors influencing PS success and found that beliefs explained unique variance within the model.

In seeking to measure students' beliefs about mathematical PS specifically, Kloosterman and Stage (1992) created the Indiana Mathematics Beliefs Scales (IMBS). The IMBS has six total scales including five original scales and a sixth scale that was adapted from questions from the Fennema-Sherman Usefulness Scale (Fennema & Sherman, 1976). Each scale was constructed to be used independently from the rest (Kloosterman & Stage, 1992). Since their creation, the scales have been widely used and applied in multiple contexts and countries (Iannone & Simpson, 2019). The present study was guided by the following research question:

Can scales 1, 5, and 6 of the IMBS be reduced in length to reduce survey fatigue while retaining adequate internal consistency?

Survey fatigue can be described as over-exposure due to survey length; effort required to respond or repeat administration; this can result in participants feeling overwhelmed and lead to potentially incomplete or lower fidelity data, or even participant withdrawal from studies (Fass-Holmes, 2022).

Methodology

Participants

The participants consisted of 227 middle-school students with 121 in 6th grade and 106 in 7th grade. All participants were drawn from a large, suburban, district that was located on the West Coast of the United States. Participants' ages ranged from 11.5 years old to 14 (M = 12.43, SD = 0.57) with 116 identifying as male, 111 as female, four as non-binary, and one who preferred not to self-identify gender. Regarding ethnic identity, 24 identified as Black or African American, six as Asian, 79 as Latin(x), one as Native American, 65 as White Non-Latin(x), 30 identified as "Other", 10 as two or more races, and 12 who preferred not to self-identify.

Data Sources

We administered three scales that were selected based on the goals of the cooperating district. Specifically, data were collected on IMBS belief scale 1 (*I can solve time-consuming mathematics problems*), belief scale 5 (*Effort can increase mathematical ability*), and belief scale 6 (*Mathematics is useful in daily life*). Each scale consists of 6 Likert-style items for a total of 18 questions. The scales were designed for use with secondary and college students and evidence of validity for the scales include item structure through item scale correlations and test content through expert review, in addition to reliability calculations (Kloosterman & Stage, 1992; Krupa et al., 2024). All items were administered using a continuous sliding scale allowing any value between 0 and 100, inclusive, to improve the reliability of the measures (Schraw, 2009).

Analysis

All analyses adhered to methods from Tabachnick and Fidell (2019). All data were tested for requisite statistical assumptions prior to data analysis, including univariate and multivariate normality, collinearity, reproducibility of the correlation matrix, univariate and multivariate outliers, and the Kaiser-Meyer-Olkin (KMO) Test of Sampling Adequacy (Tabachnick & Fidell, 2019). Data were normally distributed at the univariate (all skewness and kurtosis values were less than the absolute value of 2) and multivariate levels (all standardized residuals were less than 2 standard deviations of their respective means), with no collinearity present in the data (all zero-order correlations were < 0.80). Further, outlier analyses revealed no extreme outliers at the univariate (via box-and-whisker plots) or multivariate level (via Mahalanobis Distance).

Descriptive statistics were computed for all measures utilizing IBM SPSS 27 software. Exploratory factor analysis (EFA) with common factor extraction (principal axis factoring [PAF]) and oblique rotations (promax) were employed to examine whether the original scale could be reduced in length. We chose this approach for two reasons. First, our analyses were grounded in theoretical assumptions regarding the relations among these indicators of mathematics anxiety, and hence, justifying the EFA rather than the principal components analysis (PCA), which is atheoretical and purely statistical.

Second, we selected PAF as our extraction method because, unlike PCA, which assumes all communalities to be 1, PAF employs the multiple squared correlation coefficient, R^2, to determine communalities after extraction. Also, unlike maximum likelihood extraction, which attempts to maximize the variance of the solution and may overestimate the explained variance, PAF is a more conservative solution.

Finally, we employed an oblique rotation because we assumed, based on theoretical considerations, that the factors, if multiple, would, in fact, be correlated. The overall model fit, the standardized factor loadings, and the explained variance each factor contributed to its indicators were analyzed for this purpose for the reduced version of the measure. Our modeling procedure began by including all 18 of the original items. We chose standardized factor loadings ≥ 0.35 because, as a measure of effect, this indicates that ~12% of the item's variability is attributable to the latent variable (Tabacknick & Fidell, 2019).

Results

Descriptive statistics, internal consistency reliability coefficients (Cronbach's alpha), and the zero-order correlation matrix for the original IMBS scale and for the shortened 11-item scale are presented in Table 1.

Table 1

Descriptive Statistics and Zero-Order Correlation Matrix for the Three IMBS's for Mathematical PS Beliefs for the Original 18-Item Measure and the Shortened 11-Item Version

Scale	*M*	*SD*	α	1	2	3
1. Difficult Problems ([original] 6/[shortened] 5 items)	78.84 [a]/ 75.53 [b]	17.48 [a]/ 17.53 [b]	.87 [a]/ .85 [b]	-	.47*	-.41*
2. Math Utility Value ([original] 6/[shortened] 3 items)	58.52 [a]/ 32.33 [b]	20.74 [a]/ 5.72 [b]	.66 [a]/ .76 [b]	-.41**	-	-.40**
3. Effort ([original] 6/[shortened] 3 items)	38.13 [a]/ 56.59 [b]	4.18 [a]/ 22.42 [b]	.69 [a]/ .77 [b]	.35**	-.35**	-

* $p < .05$ ** $p < .01$ (one-tailed test of significance) *Note*. The correlations above the diagonal are for the original 18-item measure and those below the diagonal are for the shortened 11-item measure. [a] Original 18-item measure [b] Shortened 11-item measure; N = 227

Factor Analyses

The EFA results with common factor extraction–PAF–and an oblique rotation (promax) were interpreted next. Inspection of preliminary analyses revealed no difficulties in the data to reproduce a correlation matrix. Finally, the KMO Tests of Sampling Adequacy was appropriate for both original scale (KMO = .864, χ^2 (153) = 1702.09, $p < .001$) and the 11-item shortened version (KMO = .886, χ^2 (91) = 1158.86, $p < .001$), thereby permitting the factor analysis to be conducted. As with the original scale, we hypothesized a three-factor solution. This decision was made for theoretical reasons and based on prior research rather than allow a freely estimated solution with eigenvalues greater than 1.

Original Indiana Scale

The EFA with a PAF common extraction and a promax oblique rotation for the original 18-item IMBS yielded a three-factor solution which explained 46.09% of cumulative variance. The correlations among the three factors ranged from $r = .12$ to $r = .64$ in absolute value. Descriptive statistics, communalities after extraction, and standardized factor loadings for this solution are presented in Table 2. One item in the word problems scale and two items in the effort scale did not load onto any factor, and hence, only 15-items of the 18 were retained.

Table 2

Descriptive Statistics, Communalities, and Standardized Factor Loadings of the Final Model for the Original 18-Item Indiana Scales for Mathematical PS Beliefs

Item	*M*	*SD*	Com.	DP	MUV	EF
DP1	81.15	21.45	.73	.90		
DP2	77.62	22.65	.59	.84		
DP3	77.00	22.29	.52	.78		
DP4*	81.53	21.42	.62	.73		
DP5*	80.27	22.65	.50	.71		
DP6*	73.67	25.56	.35	.63		
MUV1*	61.29	8.93	.59		.93	
MUV2*	64.36	12.85	.40		.77	
MUV3*	62.85	8.36	.50		.73	
MUV4	69.67	25.87	.56		.41	
MUV5	70.54	26.29	.73		.41	
EF1	58.80	28.47	.47			.73
EF2	49.30	13.03	.31			.69
EF3	63.12	29.94	.30			.51
EF4	64.44	30.11	.59			.40

Key: Com. = Communality after extraction; DP = Difficult Problems; MUT = Math Utility Value; EF = Effort. * Reverse-coded item

The EFA with a PAF common extraction and a promax oblique rotation for the shortened 11-item IMBS also produced a three-factor solution which explained 52.53% cumulative variance. The correlations among the three factors ranged from $r = .29$ to $r = .51$ in absolute value. Descriptive statistics, communalities after extraction, and standardized factor loadings for this solution are presented in Table 3.

Review of both final EFA solutions yields some intriguing findings. The original 18-item IMBS is not only longer than our shortened version, but, evidently, it also leads to a degraded solution with appreciably lower explained variance. Whereas our proposed shortened 11-item version explains over 52% of variability in the items, the original 18-item version (Kloosterman & Stage, 1992) explains only approximately 46% of the variance in the items. Contrasting the standardized factor loadings for the solutions of the original version and the shortened versions leads us to conclude that factor loadings are generally higher for our proposed shortened scale, especially the lower-bound values, as some of the items in the original longer version not only manifested lower factor loadings at the lower bound, but also lower communalities after extraction. This, along with the more parsimonious measure than its original counterpart, supports our conclusion that our proposed shortened version, S-IMBS, is the better choice, especially when combined with other measures in a longer survey.

Table 3

Descriptive Statistics, Communalities, and Standardized Factor Loadings of the Final Model for the Shortened 11-Item Indiana Scales for Mathematical PS Beliefs

Item	*M*	*SD*	Com.	DP	MUV	EF
DP1	58.48	28.40	.46	.73		
DP2	63.02	26.72	.51	.66		
DP3*	67.64	31.01	.98	.69		
DP4*	66.33	28.88	.57	.68		
DP5*	80.10	22.69	.52	.41		
MUV1*	64.38	22.68	.70		.81	
MUV2*	61.41	28.84	.50		.77	
MUV3*	63.08	28.38	.61		.76	
EF1	77.18	22.52	.65			.83
EF2	81.09	21.40	.71			.68
EF3	81.31	21.36	.66			.64

Key: Com. = Communality after extraction; DP = Difficult Problems; MUV = Math Utility Value; EF = Effort. * Reverse-coded items

Discussion, Limitations and Avenues for Future Research

The present study provides evidence that Belief Scales 1, 5, and 6 can be shortened from the original 18 items to 11 items while maintaining adequate internal consistency. Moreover, the study adds to the evidence of validity supporting IMBS. Researchers should place a greater focus on ensuring that self-report measures are as short as possible, yet reliable and valid, to increase the odds that participants will yield accurate, complete data and avoid survey fatigue when completing measures. We have met this challenge by shortening a tool that measures an important psychological phenomenon, math PS beliefs.

Despite these findings, the study is still limited by the sample of students which were all drawn from a single district. Thus, future research should consider replicating the study across contexts. Moreover, as this study only focused on three of the six scales from the original IMBS, future research should explore whether the remaining three scales can also be shortened.

Acknowledgements

The research reported here was supported by the EF+Math Program of the Advanced Education Research and Development Fund (AERDF) through funds provided to CueThink. The perspectives expressed are those of the authors and do not represent views of the EF+Math Program or AERDF.

References

Chapman, O. (2015). Mathematics teachers' knowledge for teaching problem solving. *LUMAT: International Journal on Math, Science and Technology Education*, *3*(1), 19–36. https://doi.org/10.31129/lumat.v3i1.1049

Fass-Holmes, B. (2022). Survey fatigue—literature search and analysis of implications for student affairs policies and practices. Journal of Interdisciplinary Studies in Education, *11*, 56–73.

Fennema, E., & Sherman. J. (1976). *Fennema-Sherman mathematics attitudes scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males*. Wisconsin Center for Educational Research.

Iannone, P., & Simpson, A. (2019). The relation between mathematics students' discipline-based epistemological beliefs and their summative assessment preferences. International Journal of Research in Undergraduate Mathematics Education, *5*, 147-162.

Kloosterman, P., & Stage, F. K. (1992). Measuring beliefs about mathematical problem solving. *School science and mathematics*, *92*(3), 109-115.

Krupa, E. E., Bostic, J. D., Bentley, B., Folger, T., Burkett, K. E., & VM2ED community (May, 2024). *Search.* VM2ED Repository. Mathedmeasures.org.

Lester, F. K. (2013). Thoughts about research on mathematical problem-solving instruction. *The Mathematics Enthusiast, 10*, 245-278. Retrieved from https://scholarworks.umt.edu/tme

McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning (pp. 575–596). Macmillan.

Muis, K. R. (2004). Personal epistemology and mathematics: A critical review and synthesis of research. Review of Educational Research, *74*(3), 317– 377.

Rhodes, S., Bryck, R., Gutierrez de Blume, A. (2023). Exploring factors influencing success in mathematical problem solving. *Proceedings of the 44th annual conference of the North American Chapter of the International Group for the Psychology of Mathematics Education (PME-NA)*.

Schommer, M., Crouse, A., & Rhodes, N. (1992). Epistemological beliefs and mathematical text comprehension: Believing it is simple does not make it so. Journal of educational psychology, *84*(4), 435.

Schraw, G. (2009). *Measuring metacognitive judgments*. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), Handbook of metacognition in education (pp. 415–429). Routledge/Taylor & Francis Group.

Schunk, D. H. (1991). Self-efficacy and academic motivation. Educational Psychologist, *26*(3–4), 207–231.

Sun, L., & Zhang, Y. (2024). Cultural differences in mathematics education: A comparative study. Journal of Cross-Cultural Education, *15*(1), 1–15.

Tabachnick, B. G., & Fidell, L. S. (2019). Cleaning up your act: Screening data prior to analysis. *Using multivariate statistics* (7th ed., pp. 676-780). Pearson.

Voss, T., Kleickmann, T., Kunter, M., & Hachfeld, A. (2013). Mathematics teachers' beliefs. *In Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 249-271). Springer US.

INVESTIGATING DIFFERENCES IN ASSESSMENT DELIVERY FORMATS: AN ILLUSTRATION STUDY

Toni A. May
Binghamton University
tmay3@binghamton.edu

Gregory E. Stone
Metriks Amerique
gregory@metriks.com

Jonathan D. Bostic
Bowling Green State University
bosticj@bgsu.edu

Timothy Folger
Binghamton University
tfolger@binghamton.edu

Connor J. Sondergeld
Metriks Amerique
c.sondergeld@metriks.com

This study explored how mathematics problem-solving constructed-response tests compared in terms of item psychometrics when administered to eighth grade students in two different static formats: paper-pencil and computer-based. Quantitative results indicated similarity across all psychometric indices for the overall tests and at the item-level.

Our research team has developed and validated a series of paper-pencil, vertically equated, mathematical problem-solving measures for grades 3-8 called Problem Solving Measures (PSMs 3-8) and shared findings from prior validation studies (e.g., Bostic & Sondergeld, 2015; Bostic et al., 2017). Each PSM was designed to align with the Common Core State Standards for Mathematics (CCSSI, 2011). To expand past scholarships, we began a multi-year process of developing and validating new items for a computer adaptive testing environment. Bostic and colleagues (2024) outline a validation study for the computer adaptive (CAT) mathematical problem-solving measures, which we call DEAP-CAT. During validation and development, we realized that general research on the comparability of results from paper-pencil and computer-based test formats focused primarily on multiple-choice questions and results varied depending on testing contexts (Hamhuis et al., 2020). Further, there was a dearth of research comparing assessment psychometric properties between formats specifically related to mathematical problem solving. Thus, the purpose of this study was to psychometrically compare mathematical problem-solving constructed response item assessments using the same items administered to 8th grade students in paper-pencil and static computer-based formats.

Relevant Literature

Comparing Assessment Delivery Formats

Research on the effect that assessment delivery format (i.e., paper-pencil vs. computer-based) has on testing results has yielded contrasting findings based on specific contexts, including content area, grade level, and item type (Hamhuis et al., 2020; McClelland & Cuevas, 2020;

Puhan et al., 2007). A testing *mode effect* refers to "the likelihood of differential student performance due to differences in how items are presented in [paper-pencil tests] versus [computer-based tests]" (Hamhuis et al., 2020, pp. 2341-2342). At times, research has shown that students perform better on paper-pencil tests compared to computer-based (e.g., McClelland & Cuevas, 2020; VanDerHeyden et al., 2023). However, other research found no significant difference in student performance based on testing medium (e.g., Hamhuis et al., 2020; Threlfall et al., 2007). Further, research has suggested that the existence of a testing mode effect may depend on individual students' backgrounds and characteristics (Hamhuis et al., 2020).

When specifically investigating mathematical constructed-response assessments, one study showed sixth-grade students performed better when the test was delivered in paper-pencil format rather than computer-based (McClelland & Cuevas, 2020). By taking a deeper look at *how* students engaged with mathematical word problems via paper-pencil and computer-based test formats, some research has shown students use different processes (Lemmo & Mariotti, 2017). These results imply that even if student performance in the aggregate is similar across testing mediums, it may not be appropriate to make comparisons of student performance across paper-pencil and computer-based tests (VanDerHeyden et al., 2023). As such, VanDerHeyden and colleagues (2023) concluded that "reliability for the [early-childhood arithmetic test] is only established within each assessment format…but a score obtained in computer-based conditions could not be generalized to scores obtained under paper/pencil conditions and vice versa" (p. 98). While this seems to be a budding line of inquiry, in general, there is a scarcity of research comparing psychometric properties of test items (e.g., difficulty measures, standard error, reliability, fit indices) when the same items are administered in both paper-pencil and computer-based formats, particularly for mathematics problem-solving constructed response items.

Mathematical Problem Solving

Similar to our prior testing scholarship, our research team drew upon two related frameworks for mathematical problems. One frame is that a mathematical problem is a task presented to an individual such that (a) it is unclear whether a solution or how many exist and (b) the pathway to a solution is uncertain (Schoenfeld, 2011). This framing is useful but is not comprehensive for word problem research. Hence, we draw from Verschaffel and colleagues (1999) framing for mathematical word problems as tasks presented to an individual that are open, complex, and realistic. Open tasks may be solved using multiple developmentally-appropriate strategies.

Complex tasks are not readily solvable by an individual and require productive thinking. Open and complex are connected with Schoenfeld's framing of problems. Realistic word problems draw upon real-life experiences, experiential knowledge, and/or believable events. This notion of realism adds a necessary element to effectively frame word problems for our assessment. As a contrast, mathematical exercises are mutually exclusive from problems and are intended to support building an individual's efficiency with a known procedure (Kilpatrick et al., 2001).

Given these two synergistic frameworks for the CAT items and ensuing test, we chose Lesh and Zawojeski's (2007) problem-solving framework for PSM mathematical problem-solving computer adaptive test, which reflects our past test development. Problem solving is a process of "several iterative cycles of expressing, testing and revising mathematical interpretations – and of sorting out, integrating, modifying, revising, or refining clusters of mathematical concepts from various topics within and beyond mathematics" (Lesh & Zawojewski, 2007, p. 782). Problem solving is something that takes time and concentrates goal-oriented efforts on a problem (Polya 1945/2004; Schoenfeld, 2011), which differs from completing exercises.

Method

This research is part of a large federally funded, multi-year initiative to develop and validate items for use in grades 6-8 computer adaptive problem-solving tests. We drew on a design science approach (Middleton et al., 2008) due to its effectiveness in creating assessments through a cyclical process of designing, testing, evaluating, and reflecting. The current study fits into the design science approach by testing comparability findings when PSMs were administered in paper-pencil and static computer-based formats and then reflecting on results and usability.

Participants & Instrumentation

Multiple school districts from three states in the USA representing different geographical regions (i.e., Midwest, Mountain West, and Pacific), varying contexts (i.e., urban, suburban, and rural), and the uniqueness of students' gender and ethnicity were purposefully selected for the larger project. Data from 8th grade mathematics students from those states were specifically used in this study. The samples were not identical across test administrations (because testing was anonymous), but both tests were delivered in the same schools with the same classroom teachers to maintain proximal consistency. Samples for these comparisons were 656 for paper-pencil and 490 for computer-delivered. Final sample sizes ensured we met a minimal item exposure of 30 students per item to properly calibrate performance and ensure statistical performance viability.

In addition, students identified as having a special needs or accommodation (e.g., extra time or tests being read to them) were excluded, to control for this potentially confounding variable. Only fundamental demographic data including gender- and racial/ethnic-identities were gathered and presented in Table 1.

Table 1

Final Sample Student Demographic Characteristics

Student Demographics Values	**Testing Format** Paper-Pencil (*n*=656)	Computer-Based (*n*=490)
Gender		
Female	232 (35%)	145 (30%)
Male	398 (61%)	335 (67%)
Other	4 (1%)	2 (1%)
Not Reported	22 (3%)	8 (2%)
Racial/Ethnic-Identity		
American Indian/Alaskan Native/First Nations	7 (1%)	6 (1%)
Asian	9 (1%)	10 (2%)
Black or African-American	12 (2%)	7 (1%)
Hispanic/Latino-a or Spanish Origin	48 (7%)	34 (7%)
Middle Eastern or North African	1 (1%)	0 (0%)
Native Hawaiian or Pacific Islander	7 (1%)	5 (2%)
White	551 (84%)	410 (84%)
Other	12 (2%)	7 (1%)
Not Reported	9 (1%)	11 (2%)

Our team sought to develop 240 CAT items for each grade level (i.e., grades six, seven, and eight). After numerous reviews during the item development phase of the project, a total of 182 items associated with 8th grade mathematics content standards met expectations for testing with students. A sample 8th grade item addressing Number Sense CCSSM standards is provided to contextualize the word problems created for the CAT PSMs: “A chess board is made of eight rows with eight squares in each row. Each square has an area of 3 inches^2. What is the exact length for one edge of the chess board?” Similar to past paper-and-pencil PSMs, the CAT PSMs are scored dichotomously.

Data Collection and Analysis

Tests for each delivery format were created using the same bank of 203 previously calibrated problem-solving items. All items in the bank were deemed functional during previous statistical evaluations and linked to one of the five content domains within 8th grade. To ensure comparability of item calibrations both *across* delivery models (i.e., paper-pencil, computer-based) and *within* each delivery model, a common item equating process using linking items was employed (Kedlermen, 1988). Linking items represent previously calibrated items that are consistent across all versions of the test to ensure that the calibration of items and person abilities are equivalent. Further, using Rasch (1960/1980) modeling places all items on the same linear scale. Linking ensures direct comparability of results and performance statistics.

Paper-pencil tests were designed to be completed in a single 30-minute period, for ease of administration in the classroom. Forty-four versions of such tests were constructed, each consisting of one common item (used for equating) and three to four additional unique items. Each test covered at least four of the five standard domains within 8th grade. Each student, within a class period, took an identical paper-pencil test. Students were able to make use of classroom-provided calculators and scratch paper as needed during the administration.

Computer-administered static-tests were designed and delivered through the FastTest System© (Assessment Systems Corporation, 2023) online under the same conditions used for paper-pencil administration. The full bank of items was entered into the FastTest System© and a set of 44 identical tests were generated. Each test mirrored the features of the original versions. To ensure integrity between paper-pencil and computerized test versions, items that included fractions, square roots, mathematical equations, diagrams, graphs, charts, and pictures were entered into the FastTest system as JPGs. This allowed students to see the same structurally formatted item regardless of test administration format. While students could not write on their computer screen apart from typing their response in a designated response box, they were allowed to use scratch paper for their work, if desired. A classroom-supplied calculator or an electronic calculator embedded in the examination were available for students. The embedded calculator was small enough to fit in the upper corner of the screen without blocking, covering, or hiding any element of the item or its accompanying graphics. Item exposure requirements were applied and the final comparison included 11 tests common to both delivery methods.

Rasch (1960/1980) measurement for dichotomous responses was employed to conduct psychometric analyses for both research questions in this study using Winsteps software (Linacre, 2024). Rasch measurement has long shown its effectiveness in social science instrument development and validation (see Bond & Fox, 2007). Multiple psychometric indices were investigated. Rasch reliability is a measure of internal consistency (acceptable ≥ 0.70, good

≥ 0.80, excellent ≥ 0.90; Duncan et al., 2003). Separation specifies the distinct number of item or participant groups measured by the latent variable (acceptable ≥ 1.50, good ≥ 2.00, excellent ≥ 3.00; Duncan et al., 2003). Average standard error of measurement (SEM) for items provides a measure of test precision with lower values indicating greater measurement accuracy. Item infit and outfit mean-square statistics between 0.50 and 1.50 logits are most productive for measurement and anything greater than 2.00 could distort measurement (Linacre, 2002). Item point-biserial correlations must be positive in value to demonstrate they offer measurement support, while negative point-biserial correlations suggest item removal is necessary as these items contribute in opposition to the latent variable's meaning (Wright, 1992). With Rasch measurement, each item produces a difficulty measure in logits with higher values indicating an item is more challenging to answer correctly and lower values meaning an item is easier for students to correctly answer. Item difficulty measures were compared between administration formats and considered statistically similar if they fell within ±2 standard deviations.

Findings

In terms of overall test and item comparability between testing formats, all psychometrics indices were nearly identical and told the same story (see Table 2).

Table 2

Test and Item Performance Comparison

Test and Item Psychometrics	Testing Format	
	Paper-Pencil	Computer-Based
Item Reliability	0.91	0.93
Item Separation	3.63	3.57
Average Standard Error	0.62	0.76
Negative Point-Biserial Items	0 (0%)	0 (0%)
Misfitting Items	1 (2.3%)	1 (2.3%)
Statistically Easier Items	4 (9%)	2 (4.5%)

To summarize: Item reliability and separation were "Excellent" for both (Paper-Pencil = 0.91, Computer-Based = 0.93), SEM was approximately the same (Paper-Pencil = 0.62, Computer-Based = 0.76), no items had negative point-biserial correlations, and only one item was misfitting in each version. In terms of item difficulty: Among the 44 items across the 11 tests compared, 38 items (86%) performed statistically similarly (within ±2 standard deviations) regardless of the testing format. Items that differed statistically in their difficulty measure were relatively balanced with similarly small numbers being easier when delivered in paper-pencil (n = 4, 9%) and computer-based formats (n = 2, 4.5%).

Discussion and Next Steps

Our goal was to compare the consistency of test performance and the capacity of those items to measure student ability when delivered via paper-pencil or computer-based methods. While such comparisons for multiple-choice items have been widely presented in previous research (e.g., Puhan et al., 2007), a relatively small number of studies have explored constructed response options (e.g., McClelland & Cuevas, 2020), and even fewer comparisons have been made using mathematical problem-solving tests (e.g., Lemmo & Mariotti, 2017). Two notable findings were observed in our comparison study. First, no significant or practical differences were observed relative to overall test performance (e.g., Rasch reliability, separation, item statistics) when implementing in either delivery format. Second, the overall capacity for PSM items to measure persons remains largely unchanged by delivery method.

Results from this study strengthen the evidence for using PSMs and comparing results regardless of delivery mode (paper-pencil vs. static computer-based). Next steps in our work are to test the computer-based items in a computer adaptive testing (CAT) delivery mode, as part of the design science approach. Our ability to compare the outcomes associated with delivery models *before* adding the CAT component helps to ensure that any differences uncovered during this phase are not simply the result of a change in delivery format. Given that results from delivery format comparisons have widely varied (Hamhuis et al., 2020), it is critically important that anyone considering moving a test from paper-pencil to computer-based delivery build in time to test comparability of overall assessment and item psychometrics.

Acknowledgement

Ideas in this manuscript stem from grant-funded research by the National Science Foundation (NSF 2101026, 2100988). Any opinions, findings, conclusions, or recommendations expressed by the authors do not necessarily reflect the views of the National Science Foundation.

References

Bostic, J. D., & Sondergeld, T. A. (2015). Measuring sixth-grade students' problem-solving: Validating an instrument addressing the mathematics common core. *School Science and Mathematics Journal, 115*(6), 281-291.

Bostic, J. D., Sondergeld, T. A., Folger, T., & Kruse, L. (2017). PSM7 and PSM8: Validating two problem-solving measures. *Journal of Applied Measurement, 18*(2), 1-12.

Bostic, J., May, T., Matney, G., Koskey, K., Stone, G., & Folger, T. (2024, March). Computer adaptive mathematical problem-solving measure: A brief validation report. In D. Kombe & A. Wheeler (Eds.), *Proceedings of the 51st Annual Meeting of the Research Council on Mathematics Learning* (pp. 102-110). Columbia, SC.

Common Core State Standards Initiative. (2011). *Common Core State Standards for Mathematics*. http://www.corestandards.org/wp-content/uploads/Math_Standards.pdf

Duncan, P. W., Bode, R. K., Lai, S. M., & Perera, S. (2003). Rasch analysis of a new stroke-specific outcome scale: The stroke impact scale. *Archives in Physical Medicine Rehab, 84,* 950-963.

Hamhuis, E., Glas, C., & Meelissen, M. (2020). Tablet assessment in primary education: Are there performance differences between TIMSS' paper-and-pencil test and tablet test among Dutch grade-four students? *British Journal of Educational Technology*, *51*(6).

Keldermen, H. (1988). Common item equating using the loglinear Rasch model. *Journal of Educational Studies*, *13*(4), 319-336.

Lemmo, A., & Mariotti, M. A. (2017, February). From paper and pencil-to computer-based assessment: Some issues raised in the comparison. In *CERME 10*.

Lesh, R., & Zawojewski, J. (2007). Problem solving and modeling. In F.K. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning: A project of the National Council of Teachers of Mathematics*. (pp. 763-803). Charlotte, NC: Information Age.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 878.

McClelland, T., & Cuevas, J. (2020). A comparison of computer based testing and paper and pencil testing in mathematics assessment. *The Online Journal of New Horizons in Education*, *10*(2), 78-89.

Middleton, J., Gorard, S., Taylor, C., & Bannan-Ritland, B. (2008). The “compleat” design experiment. In A. Kelly, R., Lesh, & J. Baek (Eds.), *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics teaching and learning* (pp 21-46). New York, NY: Routledge.

Polya, G. (1945/2004). *How to Solve It.* Princeton, NJ: Princeton University Press.

Puhan, G., Boughton, K., & Kim, S. (2007). Examining Differences in Examinee Performance in Paper & Pencil & Computerized Testing. *Journal of Technology, Learning, & Assessment.*

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests.* (Copenhagen, Danish Institute for Educational Research), with foreward and afterword by B.D. Wright. The University of Chicago Press.

Schoenfeld, A. H. (2011). *How we think: A theory of goal-oriented decision making and its educational applications*. New York, NY: Routledge.

Threlfall, J., Pool, P., Homer, M., & Swinnerton, B. (2007). Implicit aspects of paper and pencil mathematics assessment that come to light through the use of the computer. *Educational Studies in Mathematics*, *66*, 335-348.

VanDerHeyden, A. M., Codding, R., & Solomon, B. G. (2023). Reliability of computer-based CBMs versus paper/pencil administration for fact and complex operations in mathematics. *Remedial and Special Education*, *44*(2), 91-101.

Verschaffel, L., De Corte, E., Lasure, S., Van Vaerenbergh, G., Bogaerts, H., & Ratinckx, E. (1999). Learning to solve mathematical application problems: A design experiment with fifth graders. Mathematical Thinking and Learning, 1, 195-229.

Wright, B. D. (1992). Point-biserial correlations and item fits. *Rasch Measurement Transactions, 5*(4), 174.

Innovating and Integrating: Advancing Undergraduate Education and Preservice Teacher Development in Mathematics Education

PRESERVICE TEACHERS' LEARNING ABOUT QUADRILATERALS USING AN ONLINE LEARNING PROGRESSION CURRICULUM

Michael Winer	Michael Battista	Viktoria Taroudaki
Eastern Washington University	Ohio State University	Eastern Washington University
mwiner@ewu.edu	battista.23@osu.edu	vtaroudaki@ewu.edu

We examined the use of an online, individualized, interactive Dynamic Geometry curriculum (iDGi), developed from learning progressions, with K-8 preservice teachers' (PTs) learning of prototypical defining properties of quadrilaterals. These properties express in formal geometric terms the most visually salient spatial characteristics that students use in identifying different types of quadrilaterals and the interrelationships between quadrilaterals. We tested the effectiveness of iDGi with 541 PTs in multiple courses in which they took the same online multiple-choice tests to assess their understanding before and after using iDGi. Our conceptual test analysis shows that iDGi significantly increased PTs' reasoning about quadrilaterals.

Research has found that elementary teachers' knowledge of mathematics is significantly related to their students' achievement (Ball et al., 2005). As Ball et al. (2005) argue "How well teachers know mathematics is central to their capacity to use instructional materials wisely, to assess students' progress, and to make sound judgments about presentation, emphasis, and sequencing" (p.14). Unfortunately, research has found that K-8 teachers lack the deep mathematical understanding needed to teach mathematics meaningfully (Ball et al., 2005). A major component for better preparing K-8 teachers to successfully teach mathematics is to enhance their conceptual understanding. In this project, we examined how an online learning progression (LP) based, Dynamic Geometry (DG) instructional unit supports pre-service K-8 teachers' (PTs') understanding of geometry content.

The study is an extension of a previous work that tested an online, individualized, interactive DG learning system for grades 3-10 that can be used by students independently or by teachers in classrooms (Battista, 2019). The resulting Individualized Dynamic Geometry Instruction (iDGi) learning system integrates the use of DG, formative-assessment, research-based LP, sequencing that interactively adapts to students' locations in LPs, and built-in student guidance. iDGi focuses on concepts from 2D geometry and measurement for grades 3-8.

In this project, we utilized iDGi with K-8 PTs using iDGi modules for quadrilaterals, triangles, isometries, length, and area. iDGi includes experimenter-constructed, LP-based reasoning assessments that describe how students' learning develops and progresses for these geometric topics. Prior school-based iDGi research and development supports validity of these assessments for testing impact iDGi has on student learning. For this paper, we focus on PTs' learning of properties of quadrilaterals and the following research question: *How does instruction with iDGi Quadrilateral modules and interrelated classroom instruction affect K-8 preservice teachers' content knowledge of properties of, and interrelationships between, quadrilaterals?*

Theoretical Framework

Learning Progressions

LPs are playing an increasingly important role in mathematics education. According to the National Research Council, "Learning progressions are descriptions of the successively more sophisticated ways of thinking about a topic that can follow one another as children learn about and investigate a topic" (2007, p. 214). Battista (2019) stated, "A LP for a topic (a) starts with informal, everyday-cultural, pre-instructional reasoning typically possessed by students; (b) ends with related formal mathematical concepts; and (c) indicates cognitive plateaus reached by students in moving from (a) to (b)" (p.1). Within a constructivist framework, LPs are used in assessment, standards, curriculum design, teaching, and research (Sztajn et al., 2012).

The van Hiele LP for Geometric Reasoning

The van Hiele LP, as synthesized by Clements and Battista (1992), is as follows. *Level 0 Pre–recognition*. Students are unable to identify many common shapes. *Level 1 Visual*. Students identify geometric shapes according to their appearance, but do not explicitly attend to geometric properties. *Level 2 Descriptive/analytic*. Students recognize and can characterize shapes by their properties, but do not see relationships between classes of shapes. *Level 3 Abstract/relational*. Students form sufficient definitions, distinguish between necessary and sufficient sets of conditions, interrelate shape categories, and sometimes provide logical arguments. *Level 4 Formal deduction*. Students establish theorems within an axiomatic system.

Dynamic Geometry Environments

In DG, static shapes are replaced by manipulable, dynamic, property-constrained screen objects. It has been claimed that DG provides "a revolutionary means for developing geometrical understanding" (Mariotti, 2001, p. 257). iDGi was created as an expansion of Battista's (2012b) *Shape Makers* DG curriculum by extending it into an individualized online format. iDGi research with students from 5th, 8th, and 10th grades shows gains in student understanding of quadrilaterals (Battista, 2019). We investigated if similar gains in achievement are made with PTs. Battista's (2012a) LP in Table 1, an elaboration of the van Hiele levels, is the conceptual framework underlying iDGi instruction and assessment. The goal of the iDGi Quadrilaterals unit is to move students to Level 2.3, and for older students into sublevels of Level 3. iDGi has inter- and intra-module sequencing supporting movement through this LP.

Table 1
Battista's (2012a) Learning Progression for Geometric Shapes

Level	Sublevel	Description
1		**Student identifies shapes as visual wholes.**
	1.1	Students incorrectly identify shapes as visual wholes.
	1.2	Students correctly identify shapes as visual wholes.
2		**Students describe parts and properties of shapes.**
	2.1	Students informally describe parts and properties of shapes.
	2.2	Students use informal and insufficient formal descriptions of shapes' properties.
	2.3	Student formally describes shapes' properties completely and correctly.
3		**Student interrelates properties and categories of shapes.**
	3.1	Students use empirical evidence to interrelate properties and categories of shapes.
	3.2	Students analyze shape construction to interrelate properties and categories of shapes.
	3.3	Students use logical inference to relate properties and understand minimal definitions.
	3.4	Students understand and adopt hierarchical classifications of shape classes.
4		**Student understands and creates formal deductive proofs.**

iDGi focuses, first, on developing prototypical defining properties, and second, property-based interrelationships between shapes (Battista et al., 2018). As an example, the prototypical defining properties of rhombuses are “all sides congruent.” This property expresses in formal geometric terms the most visually salient spatial characteristics that students use in identifying rhombuses. In the iDGi Quadrilaterals unit, PTs develop reasoning about properties of squares, rectangles, parallelograms, rhombuses, kites, and trapezoids. iDGi software monitors both students' answers and their reasoning and branches to various places in the module based on this information. Branching is implemented by locating students in the LP, then using LP to decide on subsequent instructional activities. PTs needed to score 80% to proceed to the next module.

Teacher Knowledge of Geometry

Although K-8 PTs, ideally, should have attained van Hiele level 4 reasoning by the end of high school, many researchers have argued that PTs should at minimum be reasoning at van Hiele level 3 to effectively teach geometry to K-8 students (e.g., Knight, 2006; Van der Sandt & Niewoud, 2003). However, many studies have found that the majority of K-8 PTs are reasoning lower than level 3 (e.g., Knight, 2006). Moreover, Knight (2006) wondered if teachers can teach geometry effectively if the target reasoning for their students is higher than their own level of reasoning. In the current research project, we examined how iDGi supports PTs' development of geometric content that they will teach, and we compared PT knowledge to student knowledge.

Methodology and Methods

We hypothesized that using the iDGi Quadrilaterals unit, along with subsequent class activities and class discussions related to these modules, would significantly increase PTs' knowledge of properties of quadrilaterals. We tested this hypothesis using the online multiple choice pre- and posttests that are built into the iDGi Quadrilaterals unit. We collected the following types of data: iDGi pre- and posttests, completion of iDGi modules, written homework related to iDGi coursework, and video recordings of all classes related to iDGi instruction.

Participants for this study were 580 volunteer K-8 PTs enrolled in a required 10-week course for their majors at a university in the Pacific Northwest. The purpose of this course is to help PTs develop a conceptual understanding of the geometry and measurement topics found in grades K-8. Data collection occurred from April 2018 to March 2024, with 26 courses taught using iDGi and one control group course. Data collection for six courses occurred during COVID-19 and were taught synchronously on Zoom. The remaining 21 courses were taught in a classroom.

PTs were asked to complete iDGi quadrilateral modules online for homework prior to in-class activities that addressed the covered concepts in iDGi modules. The course was taught using an inquiry-based teaching approach. A typical in-class lesson consisted of the following activities: 1) the instructor explained the directions for the class activity, 2) PTs worked collaboratively in pairs on the iDGi activity, 3) the instructor facilitated a whole class discussion related to their work. Class activities focused on iDGi module content or supplemental activities from Battista's (2012b) *Shape Makers DG* curriculum. The iDGi Quadrilaterals unit was completed in eight to ten 50-minute class periods. Courses were taught by four different instructors: Dr. W (23 courses), Dr. G (1 course), Dr. B (1 course), and Dr. N (2 courses). All four instructors implemented the same lesson plans and activities and were trained in how to use iDGi applications. The same instructor, Dr. W, taught all the synchronous Zoom courses and tried to teach the course with as much congruence to the face-to-face instruction as possible.

Results

Analysis of Mean iDGi Test Scores

We did a statistical analysis of pre- and posttest scores to see if the iDGi Quadrilaterals unit and class instruction are associated with the increased PTs' achievement. We removed PTs who were repeating the course as well as PTs who did not complete one of the two pre- or posttests. The iDGi quadrilateral pre- and posttests were identical and were each split into two parts, which were analyzed separately. Note: PTs were never told that the pre-post tests were identical, the pre-posttests locked after completion, and PTs were never shown test answers. One possible threat to the internal validity of the pretest/post test design using the same test is that posttest effects might be affected by taking the pretest. However, in the school-based iDGi project assessment design, there was a pretest/post test control group in which students studied other parts of their mathematics curriculum while the treatment group did iDGi instruction. The grades 8-10 control group consisted of 250 students (12 classes, 6 teachers). The scores were as follows: Pretest 1: M=55.87, SD=7.18, Pretest 2: M=55.53, SD=8.67, Posttest 1: M=59.22, SD=8.27, Posttest 2: M=60.13, SD=7.25. The pre-posttest relative increases were, Test 1: 5.99%, Test 2: 8.29% [e.g., (59.22-55.87/55.87) = 5.99%]. Thus, there seems to be only a small test/retest effect in comparison to the pretest/posttest iDGi group gains of Test 1: 70.94%, Test 2: 47.68%.

Quadrilaterals Test 1 measured attainment of Level 2.3. Quadrilaterals Test 2 measured attainment of Levels 3.2-3.4. We performed two statistical hypothesis tests, one for each of the pre-posttest pairs, Quadrilaterals Tests 1 and 2. The scores were as follows: Pretest 1: M=51.96, SD=17.71, Pretest 2: M=59.69, SD=19.5, Posttest 1: M=88.82, SD=11.74, Posttest 2: M=88.15, SD=12.35. The results for both pre-posttest pair differences were statistically significant ($p < .0001$), and we concluded that the posttest scores were significantly higher than pretest scores. Additional analysis found no evidence to support statistically significant different score improvement based on instructor, quarter, and modality. The mean scores of pretests 1 and 2 were much lower than the mean scores of posttests 1 and 2, and the standard deviations of the pretest scores were much higher than the standard deviations of the posttest scores. Overall, 99.6% of students had posttest scores greater than pretest scores for test 1 and 94% for test 2.

In addition, we looked at a control group (same instructor, equivalent activities) that did not use iDGi to see if there was a difference in understanding of quadrilaterals at the end of the term. Control group PTs (n=39) took the same posttest as treatment PTs (Posttest 1: M=83.01, SD=14.57, Posttest 2: M=81.54, SD=16.07). Our statistical analysis shows that iDGi treatment PTs' posttest scores were statistically higher than those of control group PTs (Test 1: p=.0183, Test 2: p=.015), suggesting that iDGi-based instruction helps students understand quadrilateral properties at least as well as instruction that does not use iDGi.

LP Analysis of Pre-/Posttests for PTs Compared to Students in Schools

In the original study, a large scale field-test of iDGi Quadrilaterals was conducted with 46 teachers and 2100 students in grades 5, 8, 9, and 10. Table 2 shows, by grade level, means of teacher means of percent of students who were at the Level 2.3 Two-thirds Benchmark (LP 2.3B), along with the percents for PTs. The Level 2.3B indicates that students used Level 2.3 or 3 reasoning for at least 2/3 of LP test tasks for 3 of 4 of the shapes—squares, rectangles, parallelograms, and rhombuses. The LP 2.3B is the most reasonable indicator of initial achievement of Level 2.3 reasoning because it indicates that most of the time, for a majority of the most common quadrilaterals, students used Level 2.3 reasoning.

Table 2

Grade Level Teacher Means for Quadrilaterals 1 Tests Using LP 2.3 {2/3} Benchmark

	Grade 5 (35 teachers, n=1524)	**Grade 8** (3 teachers, n=126)	**Grades 9-10** (8 teachers, n=450)	**PTs** (n=54) LP 2.3, 3 shapes	**PTs** (n=54) LP 2.3, 4 shapes	**PTs** (n=54) LP 2.3 {1}, 4 shapes
Pre- LP 2.3B	6%	65%	31%	44%	26%	4%
Post- LP 2.3B	35%	100%	90%	96%	87%	63%

Grade 5. The mean teacher percent of Grade 5 students who achieved the LP 2.3B on Quadrilaterals Test 1 increased from 6% on the pretest to 35% on the posttest. So, although Grade 5 students made significant progress toward attaining Level 2.3 reasoning, they still had much to learn. However, importantly, the percentage of Grade 5 students reaching the LP 2.3B on the posttest (35%) was greater than the percent of Grades 9-10 students achieving it on the pretest (31%)—a testament to the amount of progress the Grade 5 students made in iDGi. Also, on the posttest, the mean percent of Grade 5 students per teacher who used Level 2.3 reasoning for at least one task for 3 out 4 of the basic shapes increased from 50% to 78%. Using L2.3 reasoning at least once for each shape indicates transition toward Level 2.3 reasoning. So, about 3/4 of the Grade 5 students were at least in transition to Level 2.3 reasoning on the posttest.

Grade 8 (Advanced). The mean teacher percent of advanced Grade 8 students who achieved the LP 2.3B increased from 65% to 100%. Even more impressive, percent of Grade 8 students reasoning at Level 2.3 on ALL tasks for 3 of 4 basic shapes increased from 28% to 93%.

Grade 9-10. The mean teacher percent of Grade 9-10 students who achieved the LP 2.3B increased from 31% to 90%, a huge change. However, the percentage of students' reasoning at

Level 2.3 on ALL tasks for 3 of 4 basic shapes increased from 7% to only 56%, evidence of the difficulty that even high school students have using Level 2.3 reasoning consistently.

PTs. The mean percent of PTs who achieved the LP 2.3B increased from 44% to 96%. The 44% of PTs is far below the post-percent for Grade 9-10 students—suggesting that often PTs understand quadrilaterals less well than iDGi-instructed high school students. Although, the percent of Grade 9-10 students' reasoning at Level 2.3 on ALL tasks for 3 of 4 basic shapes increased to only 56%, the percent of PTs' reasoning at Level 2.3 on ALL tasks for 3 of 4 basic shapes increased from 26% to 87%, further evidence of the effectiveness of iDGi in moving PTs to Level 2.3 reasoning. The percent of PTs reasoning at Level 2.3 on ALL tasks for ALL 4 basic shapes increased from 4% to 63%, indicating the difficulty PTs had in *always* using Level 2.3.

LP Analysis for *Quadrilaterals 2* Pre-/Posttests: PTs and Grade 9-10 Students

Table 3 shows, for *Quadrilaterals Test 2*, the mean teacher percentages of iDGi students in each grade level who were at the Level 3 70% Benchmark, indicating that students used Level 3 reasoning for at least 7 out of 10 of the LP test tasks on Quadrilaterals Test 2. For grade 8, attainment of Level 3 reasoning increased from about half the students to almost all the students. For grades 9-10, attainment of Level 3 increased from about one-fifth of the students to almost two-thirds of the students. And for PTs, attainment of Level 3 increased from 41% to 94%, suggesting the effectiveness of iDGi in moving PTs toward Level 3.

Table 3

Teacher Means for Quadrilaterals 2 Tests Using LP 3 70% Benchmark

	Grade 8	Grades 9-10	PTs
Pre- LP 3 {70%}	49%	19%	41%
Post-LP 3 {70%}	94%	67%	94%

Conclusions and Discussion

Research suggests that over a variety of curricula in a variety of countries and using a variety of assessments, a reasonable estimate for the percentage of students who achieve (on posttests) Level 2 or higher reasoning in the van Hiele LP in any of grades 5-9 is about 36% (Battista, 2012a; Clements & Battista, 1992). Furthermore, only about 60% of high school students achieve Level 2 reasoning by the end of high school geometry (Senk, 1989). We can compare the 36% Level 2 achievement reported above to the posttest means for iDGi Quadrilaterals LP 2.3B, which are 35% for Grade 5, and 100% for bright Grade 8 students. Also, compared to the 60% of high school students reaching Level 2 at the END of high school geometry, we have 90% for iDGi grades 9-10 students and 96% for PTs. Overall, then, iDGi students and PTs showed impressive gains in developing Level 2.3 reasoning. One *might* argue that K-8 PTs who used iDGi have sufficient attainment of LP 2.3 to teach properties of quadrilaterals to Grade 5 students, but they may struggle with advanced Grade 8 students on interrelationships.

This research adds to the discussion of how teacher educators should design content courses based on LP to support PTs' learning of K-8 geometry. These findings encourage other researchers to develop similar curricula for non-geometric topics. More research is needed on using LP tests to measure PT's content knowledge and inform instruction in PTs' content courses. Finally, research should investigate impact of iDGi-taught teachers on student learning.

References

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)*,* 14-46.

Battista, M.T. (2012a). *Cognition-based assessment & teaching of geometric shapes: Building on students' reasoning*. Portsmouth, NH: Heinemann.

Battista, M. T. (2012b). *Shape makers: Developing geometric reasoning in the middle school with geometer's sketchpad version 5*. Emoryville, CA: Key Curriculum Press.

Battista, M. T., Frazee, L. & Winer, M. (2018). Investigating How Spatial Reasoning Is Used by Middle School Students Working in an Online Dynamic Geometry Program [Isometries, Length, Angle, Area, Volume]. In Mix, K. S. & Battista, M. T. [editors]. *Visualizing Mathematics: The Role of Spatial Reasoning in Mathematical Thought*. Springer.

Clements, D.H., & Battista, M. T. (1992) Geometry and spatial reasoning. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics*. (pp. 420-464). New York: Macmillan.

Knight, K. C. (2006) *An investigation in the change in the van Hiele levels of understanding geometry of preservice elementary and secondary mathematics teachers*. Unpublished Master's Thesis. University of Maine.

Mariotti, M. A. (2001). Justifying and proving in the cabri environment. *International Journal of Computers for Mathematical Learning*, *6*, 257-281.

National Research Council. (2007). Taking Science to School: Learning and Teaching Science in Grades K-8. Committee on Science Learning.

Senk, S. L. (1989). Van hiele levels and achievement in writing geometry proofs. *Journal for Research in Mathematics Education, 20*(3)*,* 309-321.

Sztajn, P., Confrey, J., Wilson, P. H., & Edgington, C. (2012). Learning trajectory based instruction: Toward a theory of teaching. Educational Researcher, *41*(5), 147–156.

Van der Sandt, S., & Nieuwoudt, H. D. (2003). Grade 7 teachers' and prospective teachers' content knowledge of geometry. *South African Journal of Education, 23*(3)*,* 199-205.

PROSPECTIVE TEACHERS AND AI-GENERATED IMAGERY: TOWARD PEDAGOGICALLY TRANSFORMATIVE AI-USAGE

Hamilton L. Hardison
Texas State University
hhardison@txstate.edu

Carlos Nicolas Gómez Marchant
The University of Texas at Austin
nico.gomez@utexas.edu

AI is a dominant discursive theme in mathematics education. There is much talk about using AI to expedite the status quo and not enough talk about using AI in pedagogically transformative capacities. Positioning preservice teachers (PSTs) as agents of curricular change, we offer an initial report from a project exploring ways PSTs might use AI transformatively. In a number-and-operation content course, we introduced 50 PSTs to Adobe Firefly, a text-to-image AI. We present (a) Firefly-generated images and related tasks, (b) sample responses from PSTs, and (c) preliminary analyses on a task. Theory and implications are considered.

Artificial intelligence (AI) is an increasingly popular discussion topic in everyday and academic conversations, including within the field of mathematics education. Recommendations from professional organizations are emerging on the topic; for example, a recent National Council of Teachers of Mathematics (NCTM) President's Message tasked mathematics educators "to learn how to integrate [AI] into [their] instruction and into our profession" (Dykema, 2023; see also AMTE, 2022). Precisely how mathematics teacher educators should do this is presently unclear. Here, we offer at least a partial possible response to Dykema's call.

Brief Review of AI Literature

Walkington (2024) offered some progress on Dykema's task by positing four uses of AI in mathematics education and providing numerous examples of research within each category: (a) mathematics problem-solving (e.g., OpenAI, 2024), (b) mathematics tutoring and feedback (e.g., Bastani et al., 2024), (c) adapting tasks to learner needs (Norberg et al., 2024), and (d) supporting mathematics teachers in planning (e.g., Beauchamp & Walkington, 2024). These are potentially overlapping categories, and we appreciate Walkington's categorization, especially because organizational frameworks for AI in mathematics education are scarce, while uses are myriad. However, absent from Walkington's explicit categories, though perhaps not from all studies she mentioned, was attention to the extent to which AI usage was pedagogically transformative.

What is Pedagogically Transformative AI Usage?

By pedagogically transformative AI-usage, we mean new pedagogical possibilities afforded by these technological advances that would not otherwise be possible or practical. AI usage is pedagogically transformative if teachers and students interact around AI-dependent materials in new ways for a given classroom community. The notion of interaction is critical in our work (see also Steffe & Thompson, 2000). We are opposed to efforts aimed at reducing interaction between teachers and students; we have no interest in AI as a teacher replacement (cf. Erlwanger, 1973). Although there may be benefits to using AI to reduce time associated with the clerical work teaching, this is not our focus. We have little interest in perpetuating the status quo more expediently (i.e., pedagogically stagnant AI-usage); we are interested in complexifying pedagogy to gradually and recursively supplant broken aspects of the present-day educational system.

When AI usage is pedagogically transformative, both students and teachers are "in-the-loop" when AI is involved. Teachers using AI without students in mind is not pedagogically transformative. Students using AI without teacher guidance or supervision is not pedagogically transformative. This is not to suggest students or teachers cannot work with an AI platform individually. Instead, we argue teachers' anticipations and adaptations regarding students and AI usage is critical. Simply put, teachers should be anticipating conversations, experiences, and obstacles that students might have when engaging directly with AI or indirectly via AI-generated artifacts, and teachers should be adaptive when comparing their anticipations and actual enactments. Thus, the job of the teacher includes holding students, discourse, and AI-dependent didactic objects in mind. Further foundations for our work are offered below.

Toward Theoretical Foundations for Our Work

A complete exposition of the theoretical foundations of our work is not possible here. For brevity, we link this work to Greenstein and colleague's (2023) work positioning teachers as agents of curricular change and Thompson's (2002) theory of didactic objects. In Greenstein and colleague's work, PSTs were tasked to design and 3D-print manipulatives that might be useful for supporting learning about mathematical ideas. Thus, teachers are positioned as curricular makers in addition to curricular interpreters. Like Greenstein et al. (2023), we position teachers as agents of curricular change and are interested in new pedagogical possibilities afforded by new technologies. Different from Greenstein and colleagues who considered 3D-printed curricular materials designed by PSTs (and rationales for them), we consider digital curricular materials created by PSTs via AI (and rationales for these materials).

From Thompson (2002), a curricular material (e.g., a drawn representation, a 3D-printed manipulative, or an AI artifact) can be conceived as a didactic object, which is "a 'thing to talk about' that is designed with the intention of supporting reflective mathematical discourse" (p.

198). Thompson stipulates that didactic objects are not things in themselves; instead, didactic objects can only be called such when they are coupled with interactional considerations (e.g., anticipated hypothetical conversations about the didactic object with an eye toward particular mathematical ideas). Here, we focus on artifacts PSTs' created via a particular generative AI, Adobe Firefly and the rationales they provided for these artifacts. The coupling of artifacts and rationales provides evidence for classroom interactions PSTs anticipated when they created the artifacts. Therefore, these couplings function as didactic objects.

Context, Participants, and Methods

We discuss our first-time incorporation of Adobe Firefly, a text-to-image generative AI web application, into a math content course for PSTs. Firefly is described on Adobe's website as:

> Adobe Firefly is a family of creative generative AI models. Features powered by Firefly are embedded in Adobe's flagship apps and Adobe Stock. The vision for Adobe Firefly is to help people expand upon their natural creativity. As both a standalone website and a technology that powers features inside Adobe apps, Firefly offers generative AI tools made specifically for creative needs, use cases, and workflows. (Adobe, Jan. 2025)

At the time of the study, Adobe had released its second Firefly Image Model, which was advertised to be more advanced than the first. More critically, Adobe Firefly had also just transitioned from an entirely free firefly platform to a two-tiered credits-based platform. Users at the free-tier received a smaller number of credits per month (25) than did the pay-tier (1000). At the first author's request, the university provided all participating PSTs access to the pay tier. For further information about Firefly, see Gómez Marchant & Hardison (2024).

The first author taught two sections of a number-and-operation content course for PSTs who were seeking teacher certification for elementary or middle grades in Spring 2024. Across both sections, 50 PSTs participated in the study. Here, we discuss the introduction of Firefly, as well as implementation, and results from one Firefly-dependent activity. Using data from the 40 participants who completed this assignment, we present examples and a preliminary analysis of PSTs responses. We note that all PSTs indicated having no prior experience with Firefly.

Firefly-Dependent Didactic Objects and Rationales

Introduction to Adobe Firefly

Firefly produces images based on text descriptions (i.e., prompts) entered by users. However, initial image results may not fit users' expectations. For example, on Day 1 of the course, the instructor introduced PSTs to Firefly by prompting, “high school math classroom.” Firefly produced a set of four images, one of which is shown in Figure 1 (Left). Notably absent from the first-draft images were students and teachers. Then, the instructor crafted a second-draft prompt, “high school math classroom with lots of students talking and laughing, computers, skylight, windows, trees” to generate more desirable (to the instructor) images; one of these second-draft images is shown in Figure 1 (Right). Through this first-day in-class demonstration, the instructor offered PSTs a strategy for achieving incremental progress in Firefly images by reflecting on previous image outputs and varying subsequent text inputs.

Figure 1

Instructor-Engineered Images, First (Left) and Second (Right) Drafts

A Reflection on Addition/Subtraction and Adobe Firefly

After having discussed problem types for addition/subtraction (Carpenter et al., 1999), PSTs’ were tasked with a reflection that involved (a) writing a compare-difference-unknown (CDU) story problem, (b) crafting and refining prompts to produce a Firefly image with countable items matching those in their story problem, (c) selecting a final image suiting their problem, and (d) writing a paragraph to explain why they designed the image the way they did. A compare difference unknown problem involves two finite disjoint sets of known sizes; the unknown to be determined is the difference in the sets’ sizes. Because Firefly does not handle numerical specifications precisely, we also provided PSTs’ written and in-class instructions on how to use Firefly’s “Edit” features; this way, PSTs could remove or add items to a generated image as desired. Sample problems, prompts, and images appear in Figure 2. For a sample explanation, see Figure 3.

Preliminary Analyses

Forty PSTs’ authored submissions for this assignment. Our analyses target the four separate components: problems, prompts, images, and explanations. Each of these components provides insights into PSTs' anticipations of using the image in a hypothetical mathematically pedagogical interaction. For problems and prompts, we computed basic descriptive statistics regarding number choices and character counts. For images and explanations, we leveraged open and axial coding (Corbin & Strauss, 1990) to identify common features of the didactic objects PSTs crafted. Currently, our analysis incorporates 4 image codes and 12 explanation codes (5 codes reflecting edits PSTs made and 7 codes reflecting features mentioned; see Figure 4). The rationale accompanying the candies image in Figure 3 received the following explanation codes: *students solving task, simple background, multiple ways to use image*, and *attention to item color*.

Figure 2

Problems, Prompts, and Images Produced by PSTs

Problem-Prompt-Image A	Problem-Prompt-Image B	Problem-Prompt-Image C	Problem-Prompt-Image D	Problem-Prompt-Image E	Problem-Prompt-Image F
Problem: Emily has 4 soccer balls. Sarah has 2 soccer balls. How many more soccer balls does Emily have than Sarah?	**Problem:** Mary brought 9 small candies and Jack brought 5 big candies. How many more candies did Mary being than Jack?	**Problem:** Sarah has 9 yellow dinosaurs and 5 pink dinosaurs. How many more yellow dinosaurs does she have than pink dinosaurs?	**Problem:** Ms. Johnson has 5 students in her class. Mr. Hernandez has 8 students in his class. How many more students does Mr. Hernandez have than Ms. Johnson?	**Problem:** Tom has 5 apples and Tim has 3 apples. How many more apples does Tom have than Tim?	**Problem:** Max has 5 chocolate donuts. Ellie has 12 chocolate donuts. How many more chocolate donuts does Ellie have than Max?
Prompt: Two stacks of soccer balls next to each other, simple background	**Prompt:** 5 blue candies and 12 pink candies on a plate	**Prompt:** 6 yellow toy dinosaurs and 7 pink toy dinosaurs on a bedroom floor, birds eye view	**Prompt:** Split picture, two different indoor classrooms with lots of desks	**Prompt:** [None provided]	**Prompt:** Box of 12 chocolate donuts

Figure 3

A PSTs Explanation for Problem-Prompt-Image B

I designed this image to look like this because there are multiple ways to look at and solve this problem, including different sizes or colors. I thought this would create a good situation to understand students' thinking and possible points of error (like counting color instead of size, etc.) I edited the background to make sure it was pretty simple so counting would be clear and not cause confusion. I was pleasantly surprised that the candies were clearly able to be counted on the one of the first tries generating. With a little bit of editing I ended up with a good picture that fit my problem pretty quickly.

Preliminary Results for Reflection 3

Of the 40 submissions, six included non-CDU problems. Subsequent results reflect the 34 PSTs who wrote CDU problems. In the problems PSTs authored, set sizes tended to be small (ranged from 1 to 76 with mean 6.8 and median 4) as did unknown differences (ranged from 1 to 11 with mean 2.8 and median 2). Prompts were submitted by 24 of the 34 PSTs. Prompts tended to be short with an average of 68 characters. For comparison, problems tended to be about twice as long with an average of 137 characters.

Figure 4

Current Explanation Codes

Image Codes

- *Literal Collection* – Each item in distinct sets A and B are visually represented in appropriate (story-based) quantities; no items belong to both sets.
- *Figurative Collection* – some items from set A or set B are visually represented, but not necessarily in appropriate quantities; items may belong to both sets.
- *Visually disjoint sets* –a boundary can be drawn connecting opposite image-edges such that all items in set A are on one side of the boundary and all set B's items on the other side.
- Color-differentiated sets – different colors distinguish different sets.

Explanation Codes

- Nature of post-image-generation edits: Add items, remove items, background edit, color edit, other edit.
- PT considerations: students solving task, indication of AI counting difficulties, simple background, multiple ways to use image, organization of items, attention to item color.

Images tended to feature literal collections (91%) rather than figurative collections (9%). Half (50%) of the images featured visually disjoint collections. Nearly half (44%) of the images incorporated color to differentiate sets. Our current analysis of PSTs' reflections, including features they considered and edits they reported making, appears in Tables 1 and 2.

Table 1

Explanation Codes (Number of PSTs with Percent out of 34 PSTs)

Consider Students Solving the Task	Indication of AI Counting Difficulties	Simple Background Desired	Mention Student-friendly Context	Multiple Ways to Use the Image	Discuss Item Organization	Attention to Item Color
24 (71%)	9 (26%)	7 (21%)	10 (29%)	7 (21%)	17 (50%)	19 (56%)

Table 2

Post-Generation Edits (Number of PSTs with Percent out of 34 PSTs)

PSTs Who Edited	Add Items	Remove Items	Background Edit	Color Edit	Other Edits
24 (71%)	8 (24%)	9 (26%)	3 (9%)	7 (21%)	4 (12%)

Of particular interest was that 71% of PSTs explicitly indicated considering how students would engage with the images they were producing. For example, the PST who designed Problem-Image B in Figure 2 wrote: "Although unintended, I like how the soccer balls lined up so that students could use one-to-one ideas to solve the problem." This quotation suggests the PST considered how spatial configurations within the image might support hypothetical students' solution processes, and this insight was available to us due to PSTs' interaction with

Firefly. Soccer balls are rarely “lined up” like this in the physical world. Firefly permitted artistic liberties, which may have pedagogical potential as argued by this PST.

Closing Remarks

In closing, we note that Firefly afforded us new possibilities for stimulating and understanding PSTs mathematically pedagogical ways of thinking. We emphasize that we are not currently exploring AI avenues for accomplishing the current work of teaching faster. Instead, we are interested in how teachers can use AI to teach in ways that were not previously possible; though more time consuming, the actions provide teachers with more agency in developing the discourse of their envisioned classrooms. As mathematics teacher educators, we used Firefly to generate new didactic objects (Thompson, 2002) for our in-class activities with PSTs, and we encouraged PSTs to create their own didactic objects; AI facilitated creating these didactic objects. Because our analyses are ongoing, we refrain from suggesting implications.

At this time, we are satisfied with offering some powerful examples of ways PSTs used Firefly. From our perspective, AI-generated mathematically pedagogical imagery offered a novel context for PSTs to apply, extend, or explain their mathematical thinking. We also note that PSTs’ Firefly-dependent submissions offered us multimodal insights into PSTs’ mathematical and pedagogical thinking as we examined the prompts they engineered, the Firefly images they generated and selected, and the explanations they authored to explain and reflect on their activities. In subsequent analyses, we hope to further refine codes to more robustly characterize the didactic objects PSTs created. Furthermore, we hope to complete analyses on other tasks, which will reciprocally influence the analyses presented here.

References

Association of Mathematics Teacher Educators. (2022). Position of the Association of Mathematics Teacher Educators on Technology. Retrieved March 7, 2024 from: https://amte.net/sites/amte.net/files/AMTE%20Technology%20Statement%20Oct%202022_0.pdf

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, O., & Mariman, R. (2024). Generative AI can harm learning. SSRN 4895486. http://dx.doi.org/10.2139/ssrn.4895486

Beauchamp, T., & Walkington, C. (2024). Mathematics teachers using generative AI to personalize instruction of students’ interests. AMTE Connections. https://amte.net/connections/2024/05/connections-thematic-articles-artificial-intelligence-mathematics-teacher

Carpenter, T. P., Fennema, E., Franke, M. L., & Empson, S. B. (1999). *Children’s mathematics: Cognitively guided instruction*. Heinemann.

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology, 13*(1), 3–21.

Dykema, K. (2023, November). *Math and artificial intelligence*. NCTM. https://www.nctm.org/News-and-Calendar/Messages-from-the-President/Archive/Kevin-Dykema/Math-and-Artificial-Intelligence/

Erlwanger, S. H. (1973). Benny's conceptions of rules and answers in IPI mathematics. *Journal of Children's Mathematical Behavior*, *2*(1), 212–231.

Gómez Marchant, C. N., & Hardison, H. L. In the shadows of burgeoning Colossi: The whiteness of AI in mathematics teacher education. *Connections*. https://amte.net/sites/amte.net/files/Connections%28Gomez%20Marchant%29.pdf

Greenstein, S., Akuom, D., Pomponio, E., Fernández, E., Davidson, J., Jeannotte, D., & York, T. (2023). Vignettes of Research on the Promise of Mathematical Making in Teacher Preparation (pp. 73–109). In F. Dilling, F. Pielsticker, & I. Witzke (Eds.), *Learning Mathematics in the Context of 3D Printing*: Springer Spektrum, Wiesbaden. https://doi.org/10.1007/978-3-658-38867-6_4

Norberg, K. A., Almoubayyed, H., De Ley, L., Murphy, A., Weldon, K., & Ritter, S. (2024a). Rewriting content with GPT-4 to support emerging readers in adaptive mathematics software. International Journal of Artificial Intelligence in Education, 1–40.

OpenAI (2024, Sept 12). Introducing OpenAI o1-preview. *OpenAI.* https://openai.com/index/introducing-openai-o1-preview/

Steffe, L. P., & Thompson, P. W. (2000). Interaction or intersubjectivity? A reply to Lerman. *Journal for Research in Mathematics Education*, *31*(2), 191–209.

Thompson, P. W. (2002). Didactic objects and didactic models in radical constructivism. In K. Gravemeijer, R. Lehrer, B. van Oers, L. Verschaffel (Eds)., Symbolizing, Modeling, and Tool Use in Mathematics Education (pp. 191–212). Dordrecht, The Netherlands: Kluwer.

Walkington, C. (2024). The implications of generative artificial intelligence for mathematics education. In K. W. Kosko, J. Caniglia, S. Courtney, M. Zolfaghari, & G. A. Morris (eds.), *Proceedings of the 46th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 2249–2265).

PRESERVICE TEACHERS' PERSONIFIED RELATIONSHIPS WITH MATHEMATICS

Tracey Howell
Appalachian State University
howellth@appstate.edu

Daniela Tîrnovan
Brooklyn College, CUNY
danielatirnovan24@gmail.com

Michael J. Bossé
Appalachian State University
bossemj@appstate.edu

This paper explores the complex formation of mathematical identity among preservice teachers, emphasizing the emotional and sociocultural influences that shape their perceptions and pedagogical approaches to mathematics. This study examines how preservice teachers' personification narratives reveal evolving relationships with mathematics as they transition from students to educators. Participants confront past struggles, reframe their relationships with mathematics, and develop student-centered approaches, contributing to more inclusive and resilient classroom environments.

Introduction

Preservice teachers' (PSTs) formation of mathematical identity is emotionally and socioculturally influenced, shaping their pedagogical approaches to mathematics. Grounded in both Cultural-Historical Activity Theory (CHAT) (Engeström, 2001; Roth & Lee, 2007) and Sfard's Identity Framework (SIF) (Sfard & Prusak, 2005), this study examines how PSTs' personification narratives reveal evolving relationships with mathematics as they transition from students to educators. Through these narratives, participants confront past struggles, reframe their relationships with mathematics, and develop student-centered approaches, contributing to more inclusive and resilient future classroom environments. This study implies that teacher education programs should consider personification activities and the importance of PSTs' mathematics identities in preparing teachers to create inclusive learning experiences for their multilingual students.

Literature Review

Understanding PSTs' development of mathematical identities can foster inclusive and resilient classroom environments. Mathematical identities are shaped by emotional experiences and broader sociocultural contexts (Di Martino & Zan, 2010; Lutovac & Kaasila, 2013; Yeh & Rubel, 2020). Societal expectations and institutional structures (Zazkis & Koichu, 2015) lead PSTs to often personify mathematics as a friend, foe, or authority. Reflective practices position PSTs to confront and reframe their relationship with mathematics, leading to empathetic, student-centered approaches in their future classrooms (Sfard & Prusak, 2005). This literature review synthesizes research on mathematical identity's emotional, cultural, and reflective dimensions.

Emotional and Sociocultural Dimensions in Mathematics Learning

Emotional and sociocultural factors shape PSTs' mathematical identities, influencing their self-concept as learners and their future classroom practices. Positive and negative emotional experiences, including anxiety, frustration, joy, and resilience, affect their evolving relationships with mathematics (Di Martino & Zan, 2010; Lutovac & Kaasila, 2013; Zazkis & Koichu, 2015). Emotions regarding mathematics are deeply intertwined with cultural and linguistic backgrounds. Multilingual students' simultaneous interpretation and translation of mathematical concepts across linguistic frameworks construct and hinder their mathematical identity (Yeh & Rubel, 2020). Academic mathematics instruction often overlooks students' language practices, alienating many multilingual students (Esmonde, 2009). However, culturally relevant and inclusive pedagogy positions language and identity to positively shape mathematical learning (Esmonde, 2009; Nasir, 2014; Yeh & Rubel, 2020). In this study, reflective narrative and personification activities act as culturally relevant and responsive pedagogical (CRP) activities and as a motivation for employing such in classrooms.

Power Dynamics and Identity Formation in Mathematics

Reflecting emotional responses, broader societal expectations, and institutional pressures, PSTs' personification of mathematics often reveals underlying power dynamics, with mathematics portrayed as a friend, foe, or authoritative exerting control over their learning. Viewing mathematics as a "challenging friend" or "trusted mentor" enables PSTs to transform previous struggles into constructive experiences, fostering resilience and growth (Sfard & Prusak, 2005; Zazkis & Koichu, 2015). However, when mathematics is personified as a "foe" or an "unapproachable authority," it amplifies anxiety and self-doubt and reinforces negative self-perceptions (Ramirez et al., 2016). Reflective practices within teacher education programs enable PSTs to confront and navigate these power dynamics, ultimately fostering empathetic and inclusive teaching approaches (Lutovac, 2019).

Reflective Practices in Mathematical Identity Formation

Reflective practices help PSTs address and reframe past mathematical experiences, fostering a positive, growth-oriented identity. Narrative and personification activities position PSTs to articulate complex emotions and transform mathematics from a "monster" or "foe" into a “supportive mentor” or “trusted friend” (Zazkis & Koichu, 2015). Such activities encourage participants to externalize and reassess previous anxieties, highlight the role of resilience, reconstruct previous fearful relationships with mathematics to empowerment and agency (Lutovac & Kaasila, 2013), share vulnerabilities, and build student-centered learning environments (Parker, 2024; Tsoli, 2023).

Theoretical Framework

This study considers how PSTs’ personification narratives reveal their evolving relationships with mathematics as they transition from learners to teachers. To explore this question, we employ two complementary theoretical frameworks, CHAT (Engeström, 2001; Roth & Lee, 2007) and SIF (Sfard & Prusak, 2005) , as robust lenses for examining the sociocultural dynamics and identity formation processes that shape teachers’ perceptions of and relationships with mathematics.

CHAT (Engeström, 2001; Roth & Lee, 2007) describes how individuals’ identities and relationships with mathematics are shaped by their participation in social, cultural, and historical contexts. This allows us to examine how PSTs’ evolving relationships with mathematics are shaped by cultural norms, expectations, and power dynamics in mathematics classrooms, teacher education programs, and practicum experiences (Strutchens et al., 2017). CHAT highlights the role of culture and language in mediating learning and identity formation (Cole & Engeström, 1993). Personification narratives allow PSTs to reflect on and, thereafter, negotiate their mathematics identities as they transition from learners to educators (Lutovac & Kaasila, 2014). CHAT recognizes existing tensions affecting change and development (Engeström, 2001). PSTs navigate the tensions between their experiences as mathematics learners and their emerging identities as mathematics educators (Strutchens et al., 2017; Lutovac & Kaasila, 2014).

SIF articulates how identities are discursively constructed through stories individuals tell about themselves and their relationships with others (Sfard & Prusak, 2005), illuminating how PSTs’ personification narratives reveal their dynamically evolving and socially mediated mathematics self-perceptions (Lutovac & Kaasila, 2014). Sfard’s distinction between “actual” and “designated” identities highlights the tension between PSTs’ current self-perceptions and the expectations they believe others hold for them (Sfard & Prusak, 2005). This distinction denotes conflicts between teachers’ current relationships with mathematics and the professional expectations of being confident, capable mathematics educators (Strutchens et al., 2017; Lutovac & Kaasila, 2014).

Integrating CHAT and SIF creates a comprehensive lens for studying PSTs' evolving relationships with mathematics through personification narratives. While CHAT provides a structured approach to examining the sociocultural dynamics and power relations within educational activity systems that shape teachers' relationships with mathematics, Sfard's framework considers the discursive construction and negotiation of mathematics identities through narrative. Together, these frameworks enable a multidimensional analysis of how PSTs' personification narratives reveal the complex interplay between their experiences, social positioning, and aspirations concerning mathematics (CHAT) and the discursive construction of identity (SIF) that influence teachers' evolving relationships with mathematics as they transition from learners to educators.

Methodology

This qualitative study employed a narrative inquiry approach to explore PSTs' evolving relationships with mathematics through the lens of personification. Personification involves attributing human qualities or characteristics to non-human entities (Paxson, 1994). Narrative inquiry explores lived experiences and how individuals make sense of their personal and professional identities (Clandinin & Connelly, 2000). Personification narratives uncover the complex interplay of personal, social, and cultural factors that shape PSTs' mathematical identities and pedagogical beliefs.

The study involved 18 undergraduate and graduate PSTs in a teacher education program at a large public university in the United States. All participants had completed at least one mathematics methods course and had some experience working with students in classroom.

Participants were asked to personify mathematics and craft a narrative that explored their evolving relationships with mathematics. The personification prompt [adapted from Zazkis and Koichu (2015)] included the following instructions:

> In about 300 words, tell a story about you and mathematics using personification (attributing human qualities or traits to a non-human entity). Consider: How long have you known each other? What does Math look and act like? How has your relationship with Math changed over time? Feel free to expand beyond these questions.

The data analysis followed a hybrid thematic approach, incorporating deductive and inductive coding strategies (Fereday & Muir-Cochrane, 2006), grounding the analysis in mathematical identity literature while remaining open to emergent themes and patterns in the participants' narratives. The data was first analyzed and coded through two overarching deductive themes from the associated literature regarding PSTs' evolving relationships with mathematics: power dynamics and sociocultural perspectives affecting their mathematical identities. The research team's iterative and collaborative analysis identified emergent inductive themes in the narratives that were not captured by the deductive codes: *Sites of Power Negotiation; Resilience, Resistance*, *and Empowerment*; *Reimagining Mathematics as a Tool for Empowerment; Cultural Contexts as Identity-Shaping Forces*; and *Sociocultural Influences on Pedagogical Beliefs*. The data was then analyzed based on these inductive themes (codes). The findings presented below are thematically grounded around these two themes with pseudonyms being used for all participants.

Findings and Discussions

Theme 1 Power Dynamics in PSTs' Evolving Relationships with Mathematics

1a) Sites of Power Negotiation

Key emotional moments, such as frustration with calculus or joy in mastering algebra, were turning points in participants' identity development. For example, Jenny described her fluctuating relationship with mathematics, from hatred to enjoyment and back to frustration, highlighting the nonlinear nature of her evolving mathematical identity. Similarly, Hanan recalled feeling anxious when encountering equations and formulas in high school, shifting her relationship with mathematics from comfort to distress. These emotional milestones, resulting from the CRP narrative and personification activities, reflect the power structures embedded in mathematical learning and the cultural norms that define success and failure in the subject.

1b) Resilience, Resistance, and Empowerment

Participants revealed how emotional challenges fostered resilience, which they carried into their teaching practice as a form of empathy for struggling learners. Participants reclaimed their sense of agency and power regarding mathematics by reframing their struggles as opportunities for growth and empathy-building. Caitlin exemplified this by stating, "Math used to make me feel stupid, but now I realize that struggling is part of the process, and I want my students to feel okay with making mistakes." This shift in perspective resulting from the personification activities overcomes power structures that often position mathematics as a gatekeeper or a source of anxiety.

1c) Reimagining Mathematics as a Tool for Empowerment

As participants envisioned themselves as teachers, they began to reimagine their relationship with mathematics, moving from a source of personal struggle to a tool for empowering students. Yarely stated, “Math felt amicable. After our many years of disagreement and a rather tiresome relationship, I finally gained an understanding of it. … I now think that I can get most students to understand more mathematics and feel the same.” By reframing mathematics as encouraging exploration and growth, Yarely positioned herself as a teacher empowered to create inclusive and supportive learning environments. This reimagining of mathematics through the lens of teacher identity shifts from mathematics as a power structure to teachers becoming agents of classroom change.

Theme 2 Sociocultural Perspectives on PSTs’ Mathematical Identities

2a) Cultural Contexts as Identity-Shaping Forces

Participants’ cultural experiences, such as learning mathematics in multilingual environments or with family support, significantly shaped their attitudes toward mathematics and teaching. Sandy shared how mathematics bridged between her and her non-English speaking parents, stating, “Because math is a universal language, my parents were able to help me even though they didn’t know English. Practicing multiplication tables on the bus with my mom helped me see mathematics as something we could share.” Similarly, Rosanny described how mathematics provided comfort and familiarity as an English language learner, noting, “My classroom became a theater of numbers, where the language barrier held no power.” These experiences illuminated by the CRP narrative and personification activities recognize diverse cultural contexts that shape PSTs’ mathematical identities and the need for culturally responsive mathematics education.

2b) Sociocultural Influences on Pedagogical Beliefs

Participants’ sociocultural experiences also informed their pedagogical beliefs regarding teaching mathematics. Jieci shared how his early struggles with memorization and word problems shaped his teaching philosophy, emphasizing the importance of supporting struggling students. Similarly, Hanan reflected on how her anxiety in high school shaped her belief in student-centered learning, stating, “I remember how hopeless I felt when I didn’t understand equations. That’s why I want to teach in a way that makes students feel seen and supported, even when math is hard." Thus, some participants reframed their struggles as opportunities to build empathy, allowing them to develop pedagogical practices prioritizing emotional well-being and academic growth.

Conclusion

This study contributes to research on mathematical identity formation by revealing how PSTs' relationships with mathematics evolve through the interplay of sociocultural factors, power dynamics, and identity formation as they transition from learners to educators. Participants' narratives highlight emotional milestones and sociocultural influences, such as cultural background, language, and family involvement, that shape their mathematical identities and pedagogical beliefs. These findings underscore the need for teacher education programs to support PSTs in building positive, agentic relationships with mathematics, fostering more inclusive, culturally responsive, and empowering mathematics learning experiences for their future students.

References

Clandinin, D. J., & Connelly, F. M. (2000). *Narrative inquiry: Experience and story in qualitative research*. Jossey-Bass.

Cole, M., & Engeström, Y. (1993). A cultural-historical approach to distributed cognition. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 1-46). Cambridge University Press.

Di Martino, P., & Zan, R. (2010). 'Me and math': Towards a definition of attitude grounded in students' narratives. *Journal of Mathematics Teacher Education, 13*(1), 27–48. https://doi.org/10.1007/s10857-009-9134-z

Engeström, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work, 14*(1), 133-156.

Esmonde, I. (2009). Ideas and identities: Supporting equity in cooperative mathematics learning. *Review of Educational Research, 79*(2), 1008–1043. https://doi.org/10.3102/0034654309332562

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods, 5*(1), 80-92.

Lutovac, S., & Kaasila, R. (2014). Pre-service teachers' future-oriented mathematical identity work. *Educational Studies in Mathematics, 85*, 129-142.

Nasir, N. S. (2014). *Racialized identities: Race and achievement among African American youth.* Stanford University Press.

Parker, G. G. (2024). *How do pre-service teachers experience becoming a primary teacher of math in the context of a mastery agenda?* (Doctoral dissertation, The Open University).

Paxson, J. J. (1994). *The poetics of personification*. Literature, Culture, Theory (6th ed.). Cambridge University Press.

Ramirez, G., Hooper, S. Y., Kersting, N. B., Ferguson, R., & Yeager, D. (2018). Teacher math anxiety relates to adolescent students' math achievement. *AERA Open, 4*(1), 1–13. https://doi.org/10.1177/2332858418756052

Roth, W. M., & Lee, Y. J. (2007). "Vygotsky's neglected legacy": Cultural-historical activity theory. *Review of educational research, 77*(2), 186-232. https://doi.org/10.3102/0034654306298273

Sfard, A., & Prusak, A. (2005). Telling identities: In search of an analytic tool for investigating learning as a culturally shaped activity. *Educational Researcher, 34*(4), 14–22. https://doi.org/10.3102/0013189X034004014

DEVELOPING MEASURES OF SELF-EFFICACY FOR TEACHING MATHEMATICS AND CONCRETE-TO-ABSTRACT PEDAGOGY

Karl W. Kosko
Kent State University
kkosko1@kent.edu

Maryam Zolfaghari
Kent State University
mzolfagh@kent.edu

Sk Rezwan Shihab
Kent State University
sshihab2@kent.edu

Ashley Aguilera
Kent State University
aaguile1@kent.edu

Enrico Gandolfi
Kent State University
egandol1@kent.edu

Richard E. Ferdig
Kent State University
rferdig@kent.edu

There are several measures of self-efficacy of teachers' mathematics or mathematics teaching. However, some scholars argue for improved validity evidence for these measures (McGee & Wang, 2014) and measures in mathematics education generally (Ing et al., in press). This paper reports initial validity evidence with K-12 preservice teachers for two different self-efficacy measures: one for mathematics teaching generally and another for a specific math pedagogy: concreteness fading. Results suggest strong initial validity evidence for these measures.

Mathematics teachers' self-efficacy for teaching is an important construct that has been found to predict classroom mathematics achievement (Küçükalioğlu & Tuluk, 2021) and has been associated with teachers' pedagogical beliefs and actions (Tillman et al., 2013). However, such results are often inconsistent throughout the literature (Davis-Langston, 2012; Küçükalioğlu & Tuluk, 2021; Tillman et al., 2013). Some scholars have suggested that despite numerous self-efficacy scales developed over the past several decades, "the creation of so many diverse measurement tools for teacher self-efficacy created confusion about the nature of self-efficacy itself" (McGee & Wang, 2014, p. 391). In addition to the critique of a lack of test content validity, many scholars have noted the general lack of validity evidence and arguments provided across measures within mathematics education (Carney et al., 2022; Ing et al., in press).

The present study reports on the construction of a validity argument for two measures of teachers' self-efficacy: self-efficacy for teaching mathematics and self-efficacy for concreteness fading. The purpose of this paper is to report the initial validity argument and accompanying evidence for these measures, with the eventual goal of assessing the affordances and constraints with self-efficacy scales of varying specificity to the domain of activity.

Theoretical Framework

Bandura (2023) defined self-efficacy as "people's judgments of their capabilities to execute courses of action that are required to attain designated types of performances" (p. 53). Following this definition, self-efficacy for mathematics teaching involves teachers' judgments of their capacity to engage in effective pedagogy to attain effective learning contexts as evidenced by students' mathematical reasoning and achievement. Unfortunately, scholars have observed mixed results regarding the relationship between teachers' self-efficacy for teaching mathematics and mean student achievement (Davis-Langston, 2012; Küçükalioğlu & Tuluk, 2021; Tillman et al., 2013). Whereas there are inconsistent findings regarding teachers' self-efficacy and their student outcomes, there are more consistent findings indicating relationships between self-efficacy scores and teachers' beliefs about and knowledge of mathematics and mathematics pedagogy (Olawale & Hendricks, 2023; Tillman et al., 2013). McGee and Wang (2014) observed this trend in research and noted that "efforts to measure teacher self-efficacy have become theoretically confused" (p. 395) with various measures conflating different forms of self-efficacy (i.e., of mathematics and of mathematics teaching), as well as belief constructs not specific towards Bandura's (2023) theory within the same measure. This led to McGee and Wang's (2014) development of a version of Tschannen-Moran and Hoy's (2001) survey specific to mathematics teaching: the *Self-Efficacy for Teaching Mathematics Inventory* (SETMI). Of note, in our own review of self-efficacy for teaching measures that are associated with higher mean student achievement, it is the Tschannen-Moran & Hoy's (2001) measure (Küçükalioğlu & Tuluk, 2021) that consistently demonstrated statistically significant and meaningful results while other common measures – e.g., Enochs et al.'s (2010) Mathematics Teaching Efficacy Beliefs Instrument (MTEBI) - did not (Davis-Langston, 2012; Tillman et al., 2013). Given this record, McGee and Wang's (2014) efforts at developing SETMI is logical, as Bandura (2023) advocated for self-efficacy measures more specific to domains of practice. Although there are adaptations of SETMI with additional validity evidence, there is relatively limited evidence for SETMI beyond psychometric analysis and limited correlational analysis (McGee & Wang, 2014).

Another potential issue with current self-efficacy for mathematics teaching measures is that they proport to assess mathematics teaching as a whole, rather than examining more tangible sub-domains of practice. Following Bandura's (2023) recommendations, we conjectured that more specified measures may be of more pragmatic use in mathematics teacher education research. To this end, we developed a measure of teachers' self-efficacy for enacting Bruner's (1966) concrete-pictorial-abstract pedagogical approach in mathematics. Now referenced as concreteness fading, this pedagogical approach involves children engaging with a mathematical concept initially with a concrete representation, then transitioning to working with a pictorial representation of the concept before finally transitioning to a fully abstracted (symbolic-only) representation (Fyfe et al., 2015; Laski et al., 2015). Thus, the Self-Efficacy for Concreteness Fading (SECF) measure was developed.

Overview and Research Question

The purpose of this paper is to present the initial validity argument for two surveys focused on self-efficacy for mathematics teaching: the SETMI, which focuses on self-efficacy for teaching mathematics in general, and the SECF which focuses on self-efficacy for teaching mathematics specifically through the concreteness fading (concrete-pictorial-abstract) pedagogical approach. We seek to examine an existing math-specific self-efficacy for teaching measure as well as a more fine-grained self-efficacy for teaching measure focused on a particular mathematics pedagogical approach. This includes review by experts for test content, cognitive interviews of preservice teachers (PSTs), and examination of data across different stages of PSTs' teacher education program. Thus, we sought to answer the following research questions:

Research Question 1 [RQ1]: *Do the SETMI and SECF assess the domains they were designed to assess?*

Research Question 2 [RQ2]: *How do SETMI and SECF distinguish between preservice teachers at different stages of their teacher education?*

Methods

Sample & Procedure

This mixed-method study aims to measure the self-efficacy level of pre-service math teachers. A convergent design (Creswell & Plano Clark, 2018) was used to merge qualitative findings from interviews with quantitative results from various data. The study was conducted at a large research-focused midwestern university in the United States. Participants included undergraduate PSTs (n = 152) majoring mainly in primary education (90.1%), as well as middle (0.7%), secondary (4.6%) or a K-12 licensure degree (4.6%). PSTs were enrolled in one of three courses: an introductory educational technology course (n = 33); an initial mathematics methods course focusing on place-value, addition/subtraction and geometry (n = 68); and a second mathematics methods course focusing on multiplicative reasoning and rational number (n = 50). Participants were predominately female (92.1%) and White (90.0%), with two participants identifying as nonbinary and several others identifying as underrepresented ethnicities (3.4% Black, 2.7% Hispanic, 2.8% Biracial/Multiethnicity, 0.7% Other Ethnicity).

All participants were asked to complete both measures within the first two weeks of the Fall 2024 semester. A subsample of 12 PSTs enrolled in the educational technology course participated in one-on-one cognitive interviews at the beginning of the semester. Prior to the beginning of the semester, an eight-member panel of experts reviewed the content validity of items. Data collection is currently ongoing, with participants being asked to complete the self-efficacy measures again at the end of this semester and in the next two semesters following.

Measures

McGee and Wang (2014) revised Tschannen-Moran and Hoy's (2001) survey by focusing on 7 of 15 items from the teaching efficacy and personal efficacy subscales and revising the wording to be mathematics specific (i.e., SETMI). SETMI demonstrated strong internal structure and reliability following a factor analysis. We used the same items from SETMI, but added math-specific adaptations of the 8 items they did not include from Tschannen-Moran & Hoy (2001). Example items from this 15-item revised SETMI (SETMI-R) are in Table 1.

Table 1

Example Items from Both Self-efficacy for Teaching Scales

	SETMI-R		SECF
5	To what extent can you adjust your math lessons to the proper level for individual students?	27	How well can you determine the activities/tasks to use manipulatives with?
8	How well can you respond to difficult mathematical questions from your students?	13	How well can you teach students to do mathematics with pictorial representations?
10	How well can you provide appropriate challenges for very capable students in mathematics?	3	How well can you introduce formal symbols/numbers when using pictorial representations?
15	To what extent can you help your students thinking conceptually about mathematics?	7	How well can you support students' mathematics learning using the concreteness fading approach?

The *Self-Efficacy for Concreteness Fading* (SECF) measure was developed based on descriptions of concreteness fading by various scholars (Bruner, 1966; Fyfe et al., 2015; Laski et al., 2015). Key in these descriptions were the focus on each particular stage being enacted in sequence. Thus, we wrote items focusing on the use of concrete manipulatives, pictorial representations, and the transition to fully symbolic representations from pictorial. We also drafted items focusing on the overarching approach to concreteness fading, using various wording. Table 1 illustrates examples of items, grouped by stage of concreteness fading.

Analysis & Results

Analysis focused on the collection and examination of validity evidence in line with recommendations from the *Standards for Educational and Psychological Testing* (AERA et al., 2014). Thus, a validity argument should integrate "various strands of evidence" (p. 21) across several studies. For this initial validity evidence (i.e., first study), we present evidence for test content and response processes to answer RQ1, and internal structure, relations to other variables, and generalizability to answer RQ2. Evidence for *test content* focuses on how well survey items assess the construct of self-efficacy for a domain. Evidence for *response processes* examines how participants engage with and interpret the items. Evidence includes cognitive interviews as well as psychometric data. Evidence *for internal structure* looks for how well items conform to a construct, and we used Rasch modeling to examine fit statistics and conduct a principal components analysis. For *relations to other variables*, we looked at correlations between the two measures as well as a t-test to examine whether PSTs further in their teacher education program had higher scores than those earlier in the program. Finally, *generalization* focuses on how well the measure can generalize to different situations and contexts, which we used item and person reliability to assess. In the pages that follow, we focus on the analysis and findings of particular data to examine the validity argument for each measure.

Expert Reviews [RQ1]

Validity evidence for test content was examined, in part, by analysis of expert review. We collected data from eight experts, including: three mathematics education researchers; three educational psychologists; and two master teachers. For each scale, experts were provided a definition of the construct to be examined, and were asked to rate each item for how relevant it was to the defined construct and provide typed notes. Following recommendations by Yusoff (2019), a content validity index (CVI) score was calculated for each item indicating how well items were rated as relevant. Acceptable CVI scores for eight experts were .83 for each item. Based on these criteria, for the SETMI-R scale, four out of 15 items were flagged for potential removal. Three were removed and the other rewritten based on expert feedback. For the SECF scale, 12 of 27 items were flagged for potential removal. Six of these were removed while the other six were revised based on written recommendations from the experts. Also, several experts noted a need for providing participants with explicit definitions for the terms concreteness fading, concrete-pictorial-abstract, and manipulatives and this was added to the survey.

Cognitive Interviews [RQ1]

Validity evidence for response processes regarding SETMI-R and SECF items was collected through cognitive interviews with 12 PSTs. Data collected was also used to examine the wording of items, specifically noting item redundancy, structure, and participant lack of knowledge on technical vocabulary. The cognitive interviews included three leading questions for each assessment item: 1) What is your answer? 2) Why did you choose that answer? and 3) What was the question asking? All six authors independently coded the transcribed interviews; across 720 responses from 12 participants, with a 99.3% agreement in coding. All SETMI-R items were interpreted by participants as intended. Three items for SECF (items 2, 11, & 27) had over 80% of participants interpreted the items as intended, but were retained for further analysis.

Quantitative Analysis [RQ1]

A Rasch analysis was conducted to assess internal reliability and provide additional evidence supporting the response processes for both the SETMI and SECF measures. Rasch converts raw ordinal data from Likert-based responses into continuous data using a logistic transformation (Bond & Fox, 2015). Rasch analysis for SETMI-R included a Rasch Principal Components Analysis of Residuals (PCAR), indicating that construct explained 56.3% of the standardized residual variance. The first contrasting construct had an eigenvalue of 1.92, suggesting the items measure unidimensional construct (Bond & Fox, 2015). The item reliability index was 0.89, with an item separation index of 2.86, indicating good differentiation across levels (strata) of items. Additionally, the mean square values for item infit (MNSQ = 0.99, Z = -0.09) and outfit (MNSQ = 1.00, Z = -0.05) confirmed that data fit the model. For the person reliability, reliability index is .93, with an item separation index of 3.61. Taken collectively, Rasch analysis for SETMI-R indicates strong evidence for internal structure, response processes, and generalizability.

For the SECF measure, we applied the same Rasch validation process as used with the SETMI-R measure. First, PCAR revealed that the SECF items accounted for 72.5% of the observed variance. Notably, the first contrast eigenvalue for SECF was 4.59, explaining 6% of the total variance, with a secondary dimension at 2.66, accounting for 3.5%. A contrast eigenvalue above 2.0 may suggest dimensionality concerns, which is further supported if disattenuated correlations fall below 0.57. However, since no such correlations were identified, any potential dimensionality in SECF appears too weak to indicate a cohesive secondary factor. The SECF measure had an item reliability of 0.92 and separation index of 3.49. Person reliability was also good with a reliability of 0.97 and separation index of 5.60. Additionally, the SECF measure showed sufficient model fit, with item infit (MNSQ = 0.98, Z = −0.16) and outfit (MNSQ = 0.98, Z = −0.21), and person infit and outfit values. As with the SETMI-R measure, results from Rasch analysis provide strong evidence for SECF's internal structure, response processes and generalizability.

Evidence for Convergent Validity [RQ2]

Convergent validity was examined with a Pearson correlation between SETMI-R and SECF scores ($r = 0.66$, $p < .001$). Results indicate a moderately strong relationship, suggesting that higher self-efficacy in teaching mathematics (i.e., SETMI-R) is associated with higher self-efficacy in applying concreteness fading (i.e., SECF). Next, we compared the self-efficacy scores of PSTs at the beginning of their first and second mathematics methods courses on SETMI-R and SECF measures. Independent samples t-tests found a significant difference in SETMI-R scores ($t = 1.85$, $p = 0.03$), with PSTs in the second course ($M = 1.65$, $SD = 1.34$) scoring higher than those in the first course ($M = 1.10$, $SD = 1.76$). For SECF, a significant difference was also observed ($t = 2.15$, $p = 0.01$), with PSTs from the second ($M = 0.56$, $SD = 2.72$) having higher scores than those from the first ($M = -0.64$, $SD = 3.17$). Results suggest that progression from the first to the second mathematics methods course is associated with increased self-efficacy in both teaching mathematics and using concreteness fading.

Discussion

Results presented here are preliminary but suggest that both the SETMI-R and SECF have good evidence towards the validity argument for each. Despite each measure being specific towards the domain of mathematics teaching and demonstrating a moderately strong correlation ($r = .66$), these measures are distinct. Results from cognitive interviews, t-tests, and Rasch analysis indicate that primary-grades PSTs tend to have lower perceived self-efficacy for concreteness fading than mathematics teaching generally. These differences may be due to what others have noted regarding specificity of self-efficacy measures (Bandura, 2023; McGee & Wang, 2014). Whether, and to what degree, such differences in scores informs the interpretability of these scores is a topic for future study. Future evidence should include longitudinal data from PSTs as well as data from in-service teachers, including data on their mathematics academic achievement for convergent validity.

References

American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME] (2014). Standards for educational and psychological testing. Washington D.C: AERA.

Bandura, A. (2023). *Social cognitive theory : An agentic perspective on human nature*. Wiley.

Bond, T., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. Routledge.

Bruner, J. S. (1966) *Toward a theory of instruction*. Harvard University Press.

Carney, M., Bostic, J., Krupa, E., & Shih, J. (2022). Interpretation and use statements for instruments in mathematics. *Journal for Research in Mathematics Education, 53*(4).

Davis-Langston, C. (2012) *Exploring relationships among teaching styles, teachers' perceptions of their self-efficacy and students' mathematics achievement* (doctoral dissertation).

Enochs, L. G., Smith, P. L., & Huinker, D. (2000). Establishing factorial validity of the mathematics teaching efficacy beliefs instrument. *School Science and Mathematics. Mathematics (SSMA)*, *100*(4), 194-202.

Fyfe, E. R., McNeil, N. M., & Borjas, S. (2015). Benefits of "concreteness fading" for children's mathematics understanding. *Learning and Instruction*, *35*, 104-120.

Ing, M., Kosko, K. W., Jong, C., & Shigh, J. C. (in press). Validity evidence of the use of quantitative measures of students in elementary math education. *SSMA.*

Küçükalioğlu, T., & Tuluk, G. (2021). The effect of mathematics teachers' self-efficacy and leadership styles on students' mathematical achievement and attitudes. Athens Journal of Education, 8(3), 221-238.

Laski, E. V., Jor'dan, J. R., Daoust, C., & Murray, A. K. (2015). What makes mathematics manipulatives effective? Lessons from cognitive science and Montessori education. *SAGE Open, 5*, 1-8.

McGee, J. R., & Wang, C. (2014). Validity-supporting evidence of the self-efficacy for teaching mathematics instrument. *Journal of Psychoeducational Assessment*, *32*(5), 390-403.

Olawale, B. E., & Hendricks, W. (2024). Mathematics teachers' self-efficacy beliefs and its relationship with teaching practices. *EURASIA Journal of Mathematics, Science and Technology Education, 20*(1), em2392.

Tillman, D. A., An, S. A., & Boren, R. L. (2013). Hispanic female preservice elementary teachers' mathematics teaching self-efficacies, attitudes, and student outcome expectations. *Journal of Mathematics Education, 6*(2), 27-47.

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and teacher education*, *17*(7), 783-805.

Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal, 11*(2), 49-54. https://doi.org/10.21315/eimj2019.11.2.6

TRANSFORMING ROLES OF PROOF: EVOLUTION OF UNDERGRADUATE STUDENTS' PERCEPTIONS

Yi-Yin Ko
Indiana State University
Winnie.Ko@indstate.edu

Abigail Craft
Indiana State University
acraft11@sycamores.indstate.edu

In this paper, we investigated how undergraduate students' perceptions of the roles of proof transformed through active-based instruction in a semester-long course. Although participants did not receive explicit instruction on de Villiers' (1990) five roles of proof, our results indicate that most students developed a greater appreciation and understanding of the roles of explanation and verification. Furthermore, our results suggest that students recognized that proof communicates mathematical concepts and facilitates the negotiation of argument validity during proof-related activities.

Introduction

Proof has gained increased attention in K-16 mathematics curricula due to its essential roles of verification, explanation, communication, discovery, and systemization, deepening learners' understanding of mathematics and enhancing the communication of mathematical ideas (e.g., de Villiers, 1990; Knuth, 2002; Stylianides, 2007). At the undergraduate level, mathematics majors are expected to construct, evaluate, and validate proofs in proof-intensive courses; however, research shows that students often struggle with proof due to overemphasizing establishing truth (e.g., Baxter-Bleiler & Pair, 2017).

As narrow perceptions of proof can lead students to overlook its necessity, Bleiler-Baxter and Pair (2017) found that five activities—presenting, discussing, conjecturing, working on problem sets, and critiquing—enhance engagement with the roles of proof as described by de Villiers (1990). Their study explored students' perceptions of these roles in an active-based proof course but did not assess alignment with de Villiers' framework, highlighting a gap in research. Understanding the alignment (or misalignment) between students' perceptions and those of professional mathematicians is crucial for identifying misconceptions about the role of proof, especially since students may retain these misconceptions despite exposure to the five roles. This paper examines undergraduate students' perceptions of the five roles of proof as well as compares these perceptions to professional mathematicians' conceptions as outlined by de Villiers (1990) in a transition-to-proof course designed to bridge non-proof-intensive courses (e.g., Calculus I, II, III) and proof-intensive courses (e.g., Abstract Algebra, Linear Algebra, and Number Theory).

Theoretical Perspective

In the mathematics community, proof serves five distinct roles—verification, explanation, communication, systematization, and discovery—highlighting the social and sense-making processes in mathematicians' proving practices (e.g., de Villiers, 1990). From a constructivist perspective, these roles are tools for validating mathematical statements and opportunities for learners to actively construct meaning and develop deeper conceptual understandings (e.g., Vygotsky, 1978). Verification focuses on establishing the truth of statements, often leading students to view proof solely to confirm correctness. However, mathematicians are interested in the explanatory power of proof, as it clarifies why a statement is universally true and promotes insights into underlying concepts (de Villiers, 1990; Knuth, 2002). Mathematicians also share ideas and organize arguments within a deductive framework of definitions, axioms, and theorems, thereby creating a systemization of results with proofs. This role of proof links seemingly unrelated concepts, fosters discovery, analyzes properties within statements, and prompts further proofs and insights. However, discovery is frequently overlooked due to the misconception that it only occurs before proof (de Villiers, 1990).

Additionally, communication is crucial in the proving process, as mathematicians share ideas to enhance the purpose of proofs in addition to the other roles (Bleiler-Baxter & Pair, 2017; de Villiers, 1990). As Hanna (1990) notes, "the acceptance of a theorem by practicing mathematicians is a social process" (p. 8). Active-based instruction of proof is essential to bridge the gap between how mathematicians and undergraduate students perceive the roles of proof. This approach fosters an environment where undergraduate students engage in proving processes similar to those of professional mathematicians (de Villiers, 1990; Cilli-Turner, 2017).

Methods

Seventeen undergraduate students enrolled in a Discrete Mathematics course—a transition-to-proof course—at a Midwest University in the United States participated in this study. The course met twice a week for 75 minutes over a 15-week semester, covering topics such as propositional and predicate logic, various proof methods (including direct proofs, proof by cases, contrapositive, contradiction, and mathematical induction), sets, relations, functions, and combinatorial methods.

The course design, created by the instructor (the first author), drew on existing proof studies and emphasized a communal method of proof instruction. This approach encouraged students to engage with proof as a social, negotiated, and sense-making process, as Ko et al. (2016) outlined. Through student-centered instruction, the instructor aimed for her students to experience and appreciate various roles of proof in mathematics, as de Villiers (1990) suggested, beyond the traditional role of verification. While fostering an appreciation for these five essential roles of proof was one of the course goals, the instructor did not explicitly name or discuss these roles during class activities. Instead, the course was designed to allow students to explore and internalize these concepts organically, without direct reference to the specific roles of proof outlined by de Villiers (1990). This implicit teaching approach aimed to provide a richer learning experience for students to explore various purposes of proof in mathematics.

The primary focus of this paper is the participants' responses regarding de Villiers' (1990) five roles of proof. Specifically, the study examines their descriptions in Assignment 1-Part 2 (pre-assessment) and Proving Journal 5 (post-assessment), with the relevant assignments detailed in Table 1.

Table 1

Questions about Five Roles of Proof in the Pre-Assessment and Post-Assessment

Pre-Assessment Assignment 1-Part 2 (Early August 2023)	Post-Assessment Proving Journal 5 (End November 2023)
Read de Villiers' (1990) article titled "The Role and Function of Proof in Mathematics." In this article, you will find five roles/functions of proof that are discussed, namely, (1) verification, (2) explanation, (3) systematization, (4) discovery, and (5) communication. After reading through the article carefully, write 2-3 complete sentences to describe each of the five roles/functions of proof in your own words. Then, you need to think back on your past learning experience with proof (either high school or college) and identify a time when you believe you were engaged in each of the five roles/functions of proof (use a different instance for each of the five functions).	I. Look back at your description of de Villiers's (1990) five roles of proof (i.e., verification, explanation, systemization, communication, and discovery) from Assignment 1-Part 2. How has your description of each role of proof changed over this semester? II. Based on your explanation in part (I) above, be specific and describe clearly and completely your recollection of the instances that led you to think differently about your description of each role of proof.

On the first day of class, a colleague requested consent to use students' written responses related to proofs for research purposes. One student declined, and two did not turn in their Proving Journal 5, resulting in data collection and analysis focused on the work of the remaining 14 students. Regarding the data analysis, an undergraduate student's conception of a specific role of proof was assigned a code of "Y" if their understanding aligned with the coding definitions. These definitions were meticulously derived from de Villiers' (1990) comprehensive description of that role, as outlined in Figure 1. Conversely, if a student's conception did not correspond to the coding definitions for that role, it was assigned a code of "N." When a student's response contained elements consistent with de Villiers' (1990) description but included inconsistent parts, we coded their work as "PA."

Figure 1

Analytical Framework for Five Roles of Proof (adapted from de Villiers' (1990) work)

Role/Function of Proof	Coding Definition	Language Used to Describe this Role/Function in de Villiers (1990)
Verification	Proof as a means to obtain conviction and establish the truth of a mathematical statement	• Verification of the correctness of mathematical statements (p. 17) • Conviction or justification (p. 17) • The idea is that proof is used mainly to remove either personal doubt and/or those of skeptics (p. 17) • Making sure (p. 17) • Concerned with the truth of a statement (p. 17)
Explanation	Proof as a means to promote understanding and illumination of why underlying mathematical concepts are true	• Why it may be true (p. 19) • Psychological satisfactory sense of illumination, i.e. an insight or understanding into how it is the consequence of other familiar results (p. 19) • Gives an understanding (p. 20) • One which makes us wiser (p. 20) • Convey an insight into why the proposition is true (p. 20)
Systematization	Proof as a means to structure unrelated definitions and previously-proved results to gain a global perspective of mathematical concepts	• The inclusion of a result in a deductive system (p. 19) • Systematization of various known results into a deductive system of axioms, definitions and theorems. (p. 20) • Intricately involved in the mathematical processes of a posteriori axiomatization and defining, … which form the backbones of both local and global systematization (p. 20) • Organize logically unrelated individual statements which are already known to be true, into "a coherent unified whole" (p. 21) • The focus falls of global rather than local illumination (p. 21)
Discovery	Proof as a means to expose unexpected results beyond the given mathematical scope or context	• New results are discovered/invented in a purely deductive manner (p. 21) • Proof can frequently lead to new results (p. 21) • To the working mathematician proof is therefore not merely a means of a posteriori verification, but often also a means of exploration, analysis, discover and invention (p. 21) • Generalize the result [to a broader class] (p. 21) • Deductive discovery via deductive generalization (p. 22)
Communication	Proof as a means to transmit mathematical thoughts and strategies to others	• A form of discourse, a means of communication among people doing mathematics (p. 22) • A human interchange based on shared meanings (p. 22) • Creates a forum for critical debate (p. 22) • The social process of reporting and disseminating mathematical knowledge in society (p. 22) • Subjective negotiation of not only the meanings of concepts concerned, but implicitly also of the criteria for an acceptable argument (p. 22)

To ensure a rigorous and comprehensive analysis, both authors independently reviewed the written interpretations provided by the participants. They then compared these interpretations using de Villiers' (1990) descriptions of the roles of proof. Following this initial coding process, both authors compared their coding. Any disagreements were discussed until the issues were resolved.

Results and Discussion

The data in Table 2 illustrate how undergraduate students' perceptions of the roles of proof in mathematics evolved over the semester. This evolution was analyzed using de Villiers' (1990) framework.

Table 2

Undergraduate Students' Perceptions on the Roles of Proof from a Mathematical Perspective in the Pre-Assessment and Post-Assessment

Roles of Proof	Alignment					
	Pre-Assessment			Post-Assessment		
	Yes (Y)	Partial (PA)	No (N)	Yes (Y)	Partial (PA)	No (N)
Verification	3 (21%)	6 (43%)	5 (36%)	2 (14%)	7 (50%)	5 (36%)
Explanation	4 (29%)	8 (57%)	2 (14%)	**6 (43%)**	6 (43%)	2 (14%)
Systematization	2 (14%)	3 (21%)	**9 (64%)**	2 (14%)	2 (14%)	**10 (71%)**
Discovery	**6 (43%)**	4 (29%)	4 (29%)	4 (29%)	3 (21%)	8 (57%)
Communication	**5 (36%)**	5 (36%)	4 (29%)	**5 (36%)**	4 (29%)	5 (36%)

Initially, students most aligned *discovery* as the primary purpose of proof, with 43% of participants (6 out of 14) viewing proof as a means of uncovering new mathematical insights. A close second was *communication*, with 36% (5 out of 14) seeing proof as a tool for expressing and critiquing mathematical ideas within their classroom community. In the beginning, *systematization* was the least recognized role, with a simplified understanding of essential organization. For students whose perceptions fully aligned with de Villiers' communication role from the start, this alignment remained stable from pre- to post-assessment. For example, Zora initially described communication as a discourse function within proof:

> Communication is the function that allows discourse within a proof. What I believe de Villiers is getting at here is that we have to understand this is going to a human audience, with a different thought process for each; the proof can be read in many lights, not just the one you mean it to be. (Pre-Assessment)

By the post-assessment, Zora's description emphasized clear and precise communication of knowledge among mathematicians:

> Proof is a tool for effective communication among mathematicians. It allows mathematicians to convey their ideas, results, and reasoning to others in a precise and unambiguous manner. Clear and rigorous proofs facilitate the dissemination and understanding of mathematical knowledge. (Post-Assessment)

While Zora's initial response focused on communication for varied audiences, her post-assessment highlighted that proof is a meaningful tool for conveying mathematical ideas. This shift reflected her growing understanding of the purpose of the communication role throughout the course.

Regarding *verification* and *explanation*, Table 2 shows that the overall percentage of students with fully or partially aligned perceptions remained constant from pre- to post-assessment. Additionally, there was notable growth among students who were initially less aligned with the explanation role. This understanding of these two roles could be tied to previous proof-related courses and the typical emphasis on the roles, especially verification (e.g., Bleiler-Baxter & Pair, 2017; de Villiers, 1990; Knuth, 2002). By the end of the semester, Table 3 demonstrates that four students shifted toward fully recognizing the explanation, discovery, and communication roles.

Table 3

Undergraduate Students' Shifts in the Roles of Proof from Pre-Assessment and Post-Assessment

Roles of Proof	Shifts from Pre-Assessment to Post-Assessment		
	Increased Shift	No Shifts	Decreased Shift
Verification	3 (21%)	8 (57%)	3 (21%)
Explanation	**4** (29%)	9 (64%)	1 (7%)
Systematization	1 (7%)	11 (78%)	2 (14%)
Discovery	**4** (29%)	4 (29%)	6 (43%)
Communication	**4** (29%)	6 (43%)	4 (29%)

These shifts occurred across various levels of alignment with de Villiers' (1990) descriptions, including transitions from no alignment to partial alignment, partial alignment to complete alignment, or no alignment to full alignment. Adam's responses illustrated this development, as he described that the explanation role "is stating that for a good proof instead of just verifying it should instead explain why the proof is a proof" in his pre-assessment. In his post-assessment, Adam commented, "[t]he explanation should go beyond just definitions and justifications but also really tell how and why the work done makes it proof." Adam's progression showed his understanding that proof is a rigorous tool for confirming truth and conveying underlying mathematical concepts. This example suggests that students developed a deeper view of the multiple purposes of proof in mathematics as they engaged more with the course materials over the semester.

As seen in Table 2, *systematization* remained the least aligned role of proof throughout the semester. Many students tend to interpret the systematization role as organizing steps, axioms, definitions, and theorems in a logical sequence within a single proof rather than as a broader structuring of unrelated definitions. They previously proved results to provide a global perspective on a mathematical concept across proofs. For example, Anna viewed systematization as breaking down and structuring mathematical concepts or conjectures in her pre- and post-assessment, reflecting a more localized understanding of the role. Similarly, Brian described systematization as "the organization of a mathematical proof" in both assessments, focusing on the arrangement of steps within individual proofs rather than on interconnections across multiple proofs. This finding suggests that undergraduate students in the course primarily viewed the systematization role as an organizational tool for individual steps rather than a method for synthesizing broader mathematical ideas. This perspective aligns with Bleiler-Baxter and Pair's (2017) findings, which indicate that many undergraduate students may have yet to fully experience systematization to link definitions, theorems, and proofs to develop a comprehensive understanding of mathematical concepts in their proof courses. Instead, it shows the breaking down of systemization as a system relating to organization, something that is expanded more in mathematics to include the global perspective of the whole, a concept nuanced enough for students who have not interacted with it previously to misunderstand (de Villiers, 1990).

Discussion and Implications

In this study, it is important to note that the undergraduate students in this active-based course did not receive explicit instruction on de Villiers' (1990) five roles of proof in mathematics until the designated week when they read and reflected on his article. Furthermore, de Villiers' (1990) roles do not encompass all perspectives on proof, nor does he claim to provide an exhaustive account of its functions. Although this study examined only 14 undergraduate students regarding the roles of proof throughout the semester, the results indicate that students' interpretations aligned most closely with the explanation and the verification roles. This alignment highlights that proof demonstrates both *how* and *why* a statement is true based on students' prior experiences with proof-based courses. Future research could explore the long-term impact of implicit instruction on college students' understanding of the roles of proof.

Although the emphasis on the communication role of proof decreased during the semester, most students generally recognized that proofs convey mathematical concepts and validate arguments during proof-related activities. This finding suggests that students may view proof as a means of communication alongside explanation and verification—an understanding critical for teaching and learning proofs (Baxter & Pair, 2017; Knuth, 2002). However, the results show that students' responses to the systematization role of proof are often misaligned with de Villiers' (1990) framework. To better enhance undergraduate students' appreciation and promote their understanding of the discovery role, future courses will incorporate tasks encouraging discovery, such as engaging students in finding unexpected results during proof construction.

References

Bleiler-Baxter, S. K., & Pair, J. D. (2017). Engaging students in roles of proof. *The Journal of Mathematical Behavior, 47,* 16-34.

Cilli-Turner, E. (2017). Impacts of inquiry pedagogy on undergraduate students conceptions of the function of proof. *The Journal of Mathematical Behavior*, *48*, 14-21.

de Villiers, M. D. (1990). The role and function of proof in mathematics. *Pythagoras, 2*4, 17-24.

Hanna, G. (1990). Some pedagogical aspects of proof. *Interchange, 2*1(1), 6-13.

Knuth, E. (2002). Secondary school mathematics teachers' conceptions of proof. *Journal for Research in Mathematics Education, 33*(5), 379-405.

Ko, Y. Y., Yee, S. P., Bleiler-Baxter, S. K., & Boyle, J. D. (2016). Empowering students' proof learning through communal engagement. *Mathematics Teacher, 109*(8), 618-624.

Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education, 38*(3), 289-321.

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

Innovating and Integrating: Collaborative Growth in Graduate and Inservice Teacher Education

GRADUATE TEACHING ASSISTANTS' BELIEFS ON PRODUCT-DRIVE PROFESSIONAL DEVELOPMENT (PRELIMINARY INVESTIGATION)

John Sevier	Rashmi Singh	Özge Gün
Appalachian State University	Appalachian State University	Bartin University
sevierjn@appstate.edu	singhr6@appstate.edu	ozgegun@bartin.edu.tr

While graduate teaching assistants (GTAs) are indispensable assets in higher education, many enter the classroom feeling underprepared to teach college mathematics courses. A case study on a cohort of GTAs focused on their engagement with product-based professional development (PD) and their beliefs on how the PD impacted their pedagogical skills and beliefs about mathematics instruction. The preliminary findings show that product creation positively impacted the GTAs' beliefs concerning PD. This research leads to the next proposed study regarding affecting GTAs' self-efficacy.

Introduction

Only recently have universities considered PD for GTAs to be an integral part of teacher education. While GTAs are indispensable assets in many undergraduate learning programs in higher education, many GTAs have backgrounds in mathematics and limited teaching experience (Di Bendetti et al., 2022). Often, they rely on their mathematical skills and self-efficacy in teaching. Many of these beliefs stem from their student experiences due to a lack of teaching experience. While pre-collegiate mathematics education literature has emphasized sound instructional practices and teacher knowledge development (Deshler et al., 2015), current PD may not align with a GTA's beliefs or self-efficacy in aligning beliefs and teaching knowledge.

In this study, we examine the research question: How does product-driven PD, in which GTAs create artifacts or lessons for their instructional use, affect GTAs' beliefs and self-efficacy about the usefulness of PD sessions? This study considers an alternative approach highlighting GTA product-based PD, wherein GTAs develop and provide evidentiary artifacts they can use in class instruction. Here, we present participating in GTAs' preliminary responses and feedback about this PD methodology.

Literature Review

Teacher Professional Development (PD)

Effective teacher PD is characterized by sustained engagement, content relevance, active learning, and collaboration (Desimone & Garet, 2015). Long-term, job-embedded PD allows teachers to reflect on their practice, integrate new knowledge, and apply instructional strategies aligned with current educational standards and the needs of diverse classrooms (Darling-Hammond, 2017).

Teacher beliefs regarding students, instructional materials, and teaching strategies develop from preservice experiences and evolve through interactions with students, colleagues, and professional learning (Noben et al., 2021). These beliefs can foster or hinder instructional innovation (Cross, 2009). For example, teachers who view mathematics as a fixed skill set may prioritize procedural teaching methods. In contrast, those with a growth-oriented perspective are more likely to adopt conceptual and exploratory approaches (Wilkins, 2008). Longitudinal studies suggest that sustained, reflective PD experiences can gradually transform beliefs, particularly when coupled with supportive school leadership and collegial collaboration (Cross, 2009). Overcoming stubborn teacher beliefs is critical for adopting reform-oriented practices (Skott, 2014).

GTAs play a vital role in higher education, supporting faculty, conducting lab sessions, and providing tutorial assistance. GTAs often enter teaching roles with limited pedagogical training, leading to variability in their instructional effectiveness (Gardner & Jones, 2011). Since the 1990s, many universities have devoted resources to GTA training programs, successfully focusing on generic teaching skills, discipline-specific strategies, and pedagogical skills (Chiu & Corrigan, 2019). These have created a better learning and teaching environment, increasing teaching competency and improving the learning experience of undergraduate students (Gardner & Jones, 2011). However, the typical duration of GTA training programs, ranging from a few days to a semester, constrains GTAs’ development. Continuous support through mentoring, peer support, the commitment of senior academics to guide and support GTAs in teaching, and in-class observation help them address ongoing teaching challenges (Deshler et al., 2015).

Previous research has reported that formal GTA training programs can boost self-efficacy in teaching and effective teaching behaviors (Chiu & Corrigan, 2019). However, extended studies on GTAs' self-efficacy beyond the initial pre-service teacher education program are limited, with mixed findings (Langdon et al., 2017).

Methodology

Three-stage Case Study Design

This project, positioned in a public state university’s mathematics department in the southeastern United States, follows a previous mixed-methods longitudinal study (Creswell &

Plano Clark, 2017) wherein the first author investigated the previous GTA cohort's perspectives on the usefulness and meaningfulness of a PD project. Based on these results and the constant turnover of GTAs, the first author felt that a future case study would be appropriate to "seek a greater understanding of the case…to appreciate the uniqueness and complexity of the case, its embeddedness and interaction with its contexts" (Stake, 2006, p. 16). Thus, we will utilize three-stage PD to follow the GTA PD experience through each phase of the GTA PD experience.

This qualitative case study examines a cohort of eight GTAs during their first term of teaching undergraduate mathematics courses as part of their departmental service. The study is structured across three distinct stages, corresponding to three academic terms of the GTAs' assistantships. PD is integrated throughout the study, with each stage incorporating a specific PD phase: Stage 1 includes phase one of PD, the Stage 2 incorporates phase two of PD, and Stage 3 implements phase three of PD. This synchronization ensures that PD aligns systematically with each stage of the case study.

Stage 1 of this three-part case study, GTAs engaged in a comprehensive 15-week preparation period that combined co-teaching with mathematics faculty and weekly two-hour PD workshops. During this initial stage, GTAs participated in dual weekly sessions while observing and co-teaching with university faculty. The GTA coordinator selected various pedagogical topics for discussion, followed by debriefing sessions on both the topics covered and co-teaching experiences. This exploratory stage was designed based on research recommendations that emphasized integrating theory with active learning, utilizing co-instructor feedback for community building, encouraging professional reflection, and implementing peer feedback for personal growth (Noben et al., 2021). The PD curriculum covered fundamental teaching elements, such as lesson planning, assignment creation, assessment development, class pacing, course planning, and classroom technology integration. This initial phase of PD served multiple purposes: to develop GTAs' understanding of mathematics instruction, enhance their confidence in course planning and daily lesson preparation, and foster a supportive community to address teaching-related anxieties and uncertainties. Additionally, this phase helped researchers identify relevant topics for the second phase of PD.

Stage 2 of the case study coincided with GTAs beginning their independent teaching assignments in their second term. The accompanying PD phase focused on three main topics: Desmos activities, asynchronous learning and lesson planning, and guided inquiry learning

(GIL). These topics, selected based on first-stage observations and feedback, were explored through in-depth, multi-week sessions that promoted ongoing discussion, reflection, and product development. Guest speakers presented each topic across three two-hour sessions spanning three weeks, during which GTAs created practical teaching artifacts like lessons or activities. Following this, GTAs completed a questionnaire to evaluate the impact of this product-oriented PD on their teaching beliefs. This stage represented study's primary data collection phase.

In the final stage, Stage 3, which aligned with GTAs' continued independent teaching responsibilities, participants selected and attended five PD sessions offered by either the university or department. GTAs documented how these sessions influenced their teaching and learning approaches. The stage concluded with a questionnaire prompting GTAs to reflect on how their second-phase PD experiences influenced both their third-phase session choices and their evolving beliefs about mathematics teaching and learning.

Data Collection

The study employed data collection through questionnaires and observations during the conclusion of the second and third stages. This methodology was selected to enable a thorough investigation of participants' viewpoints and interactions within the case study environment (Glesne, 2016). The approach provided structured yet adaptable exploration of GTA insights, allowing for the identification of themes and patterns crucial to qualitative research (Rubin & Rubin, 2012). Comprehensive, open-ended questionnaires were administered to all participants to elicit detailed feedback. The questionnaire asked GTAs to do the following: evaluate and prioritize PD sessions, provide rationales for their rankings, assess the value of creating instructional artifacts and their impact on teaching self-efficacy, share their evolving beliefs, propose future PD recommendations, and offer additional commentary regarding how the PD influenced their teaching confidence.

Observational data collection occurred during PD sessions, documenting participant interactions, behaviors, and contextual elements to enhance case understanding (Yin, 2018). Following Emerson et al.'s (2011) guidelines, systematic field notes incorporated both descriptive and reflective observations. The analysis process integrated questionnaire and observational data to identify common and contrasting themes. This dual analysis approach aimed to synthesize self-reported attitudes and beliefs with observed real-time interactions, providing a more complete understanding of GTAs' perspectives.

Data Analysis

Data analysis followed a thematic approach, with data initially coded and then categorized into preliminary themes that reflect the primary research question (Braun & Clarke, 2006). An iterative process of reading, coding, and revising codes was employed to identify patterns and emergent themes across cases after Stage 3 had concluded (Saldaña, 2016). GTA responses were coded as GTA1 through GTA8 and with a timestamp. Google software was utilized to organize and analyze data, allowing for the preliminary systematic coding and cross-referencing of themes across questionnaires. Thematic analysis enabled preliminary data exploration while preserving the complexity of participants' perspectives.

Preliminary Findings

In this section, we provide GTAs' preliminary responses and feedback about the associated PD solely from Stage 2. Full coding and theming did not take place at the time of this submission.

Questionnaire Results

All eight students responded to the questionnaire given at the end of Stage 2. Each GTA ranked their preferred second PD session (e.g., Desmos, asynchronous learning, and GIL). Most GTAs stated that the selected sessions were the most useful without additional reasoning. Others elaborated on their selections. GTA7 noted:

> I really loved learning about the different styles of teaching, like guided inquiry learning and even asynchronous teaching… I could take different pieces and try to integrate them (or at least consider them) in my classroom.

Question three asked GTAs if they believed creating artifacts or lessons/activities for their instruction was useful. Of the three answer options (Yes, No, and Uncertain), five stated uncertain, with three stating yes. Four GTAs stated that the topics "did not fit their teaching style" or "was not a good fit." GTA4 noted, "I do not think those lessons really made a difference for me or the class. I prefer the other things (previous term) to prepare for class," alluding to the Stage 1 PD. Two GTAs said they found the asynchronous sessions useful and would use them in class. GTA8 stated:

> Having to actually try some things out lets us see if we want to use them in our classroom. It's much more helpful to create the artifacts and have the experience making/using them rather than just knowing about them… That gives me more confidence that I could do that in the future, as well as opens the opportunity of being able to create this if I couldn't teach in person one day.

For question five, GTAs did not provide specific sessions/topics as suggestions but did provide potential adjustments to the current sessions/topics. Most felt the sessions could have been conducted in two weeks instead of three. Some GTAs noted adjusting the topics within the asynchronous sessions to add more discussions on humanizing the course. Conversely, two GTAs felt that more time needed to be spent on developing the asynchronous lesson or an activity.

Observation Notes

During the sessions, observations were recorded about the GTAs' involvement, emotional reactions, statements made, and reactions to questions/responses from the speakers. Observation notes for each session differed based on each topic. Positive reactions occurred for all three sessions in different phases. For the Desmos and asynchronous sessions, GTAs were more positive at the beginning of the sessions (first and second weeks) and seemed distracted, bored, and unengaged during the last session. This was corroborated by some of the statements in the questionnaire stating that sessions could have been shorter or presented on a shortened timeline.

Conversely, in the GIL sessions, GTAs seemed disengaged during the first session, but interest grew over the second and third sessions. Several mentioned that once they could see how the process worked in the classroom, they began to see the usefulness of GIL. One GTA found the GIL sessions engaging and stated, "[they] wanted to make this their teaching style." Another GTA noted they believed they could use aspects of it. One GTA adamantly disagreed with the approach, noting several times that it did not fit their teaching style and was "not applicable" to their classes.

Concerning the products, most of the GTAs' artifacts were activities they could use in their mathematics lessons, and most of the Desmos artifacts were teaching/learning activities. One GTA created a Desmos activity as an asynchronous lesson. All GTAs in the synchronous sessions produced complete lessons to implement in their class. Only four students created a GIL activity for their class. When asked, they all said they ran out of time or forgot.

Initial findings revealed mixed responses: while many GTAs valued the PD sessions and their created materials, some found limited utility in certain topics or artifacts, with one expressing concern about the practicality of the guided inquiry learning activity. Definitive conclusions about overall attitudes are premature as analysis of codes and themes is ongoing. It remains unclear whether GTAs had adequate opportunities to implement their materials. Several participants, however, plan to use these resources in future courses, which will be explored in Stage 3.

Discussion and Next Steps

The research preliminarily examined how PD focused on creating tangible teaching materials influenced GTAs' self-efficacy of such development's effectiveness. The findings could inform future mathematics teacher PD initiatives, particularly regarding program scalability and adaptability. While primarily addressing GTA development, the research also suggests opportunities for exploring how product-oriented preparation enhances GTAs' pedagogical readiness, supports student learning in mathematics, and improves overall student support systems.

The complete study aims to analyze Stages 2 and 3 comprehensively together, focusing on GTAs' developing self-efficacy regarding both the PD program and their teaching capabilities. Stage 3 provides expanded opportunities for GTAs to plan, create, modify, and implement their teaching materials. This extended implementation period aims to enable deeper self-reflection and a substantive evaluation of how PD influenced their instructional practices (Cross, 2009; Noben et al., 2021).

References

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Chiu, P. H. P., & Corrigan, P. (2019). A study of graduate teaching assistants' self-efficacy in teaching: Fits and starts in the first triennium of teaching. *Cogent Education*, *6*(1), 1579964.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.

Cross, D. I. (2009). Alignment, cohesion, and change: Examining mathematics teachers' belief structures and their influence on instructional practices. *Journal of Mathematics Teacher Education, 12*(5), 325–346.

Darling-Hammond, L. (2017). *Effective teacher professional development.* Learning Policy Institute.

Deshler, J. M., Hauk, S., & Speer, N. (2015). Professional development in teaching for mathematics graduate students. *Notices of the AMS, 62*(6), 638-643.

Desimone, L. M., & Garet, M. S. (2015). Best practices in teachers' professional development in the United States. *Psychology, Society, & Education, 7*(3), 252-263.

Di Benedetti, M., Plumb, S., & Beck, S. B. (2022). Effective use of peer teaching and self-reflection for the pedagogical training of graduate teaching assistants in engineering. *European Journal of Engineering Education*, 1-16.

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (2011). *Writing ethnographic fieldnotes* (2nd ed.). University of Chicago Press.

Gardner, G. E., & Jones, M. G. (2011). Pedagogical preparation of the science graduate teaching assistant: Challenges and implications. *Science Educator, 20*(2), 31–41.

Glesne, C. (2016). *Becoming qualitative researchers: An introduction* (5th ed.). Pearson.

Langdon, J. L., Schlote, R., Melton, B., & Tessier, D. (2017). Effectiveness of a need supportive teaching training program on the developmental change process of graduate teaching assistants' created motivational climate. *Psychology of Sport and Exercise*, *28*, 11-23.

Noben, I., Deinum, J. F., Douwes-van Ark, I. M., & Hofman, W. A. (2021). How is a professional development programme related to the development of university teachers' self-efficacy beliefs and teaching conceptions? *Studies in Educational Evaluation*, p. *68*, 100966.

Rubin, H. J., & Rubin, I. S. (2012). *Qualitative interviewing: The art of hearing data* (3rd ed.). Sage Publications.

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

Skott, J. (2014). The promises, problems, and prospects of research on teachers' beliefs. In *International Handbook of Research on Teachers' Beliefs* (pp. 13–30). Routledge.

Stake, R. E. (2006). *Multiple case study analysis*. Guilford Press.

Wilkins, J. L. (2008). The relationship among elementary teachers' content knowledge, attitudes, beliefs, and practices. *Journal of Mathematics Teacher Education*, pp. 11, 139–164.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage Publications.

INFLUENCES OF COMMUNITIES OF PRACTICE ON A MATHEMATICS-RELATED TEACHER IDENTITY

Kelly Navas
University of South Florida
knavas@usf.edu

This study explores how a mathematics-related teacher identity influences the learning opportunities teachers provide. A teacher's identity is shaped by the different interactions within various communities of practice (CoPs) at their schools. This can be visualized as interlocking gears that work together to build a teacher's identity. These experiences contribute to the ongoing development of their identity. Understanding these dynamics can help schools and preparation programs equip teachers to navigate work and communication in their CoPs. While, additionally, fostering supportive communities that encourage positive identity growth, leading to improved math instruction and more equitable student learning opportunities.

In recent decades, mathematics education research has begun to focus on the impacts an individual's identity has on student engagement, teacher instruction, and retention (Aragon, 2016; Day et al., 2005). Most people can recall pivotal moments that shaped their relationship with mathematics, whether positively or negatively (Wood, 2013). Understanding how these experiences influence teachers' decisions in the classroom is valuable, as the saying goes, "we teach who we are" (Lutovac & Kaasila, 2018, p. 760). This is particularly true for mathematics, a subject that provokes strong emotions (Hodgen & Askew, 2007; Navas, 2023). By exploring the influences of mathematics-related teacher identity, we can better support teachers in making instructional and classroom decisions that foster a positive learning environment.

The purpose of this narrative inquiry was to understand the ongoing construction of a first-year teacher's mathematics-related teacher identity. To explore this, I used a communities of practice (CoPs) lens (Lave & Wenger, 1991), examining how the teacher's identity developed through the interactions and sense of belonging within three different CoPs: grade team, classroom, and parent communities. The teacher in this study, Suzie, struggled learning mathematics and had only begun to gain positive experiences with mathematics during her teacher preparation program. Her story highlights the unique tensions she faced within each CoP impacting her planning and delivery of her mathematics lessons.

Theoretical Framework

To guide my understanding, I utilized three theoretical frameworks: the theory of experience, mathematics-related teacher identity, and sociocultural theory. Experiences are at the heart of this study, emphasizing the importance to delve deeply into their meaning. Dewey's (1997) theory of experience provides a lens for this understanding. Within this theory, there are the principles of continuity and interaction. The principle of continuity explains how every experience builds from one to another, with past experiences shaping future growth and interaction, similar to identity. The principle of interaction relates to the social dynamics between people and their surroundings (Dewey, 1997). Additionally, Gee (1999) highlights how social contexts influence individuals' ability to engage and participate within groups. Experiences form the foundation of how individuals view themselves and interpret others' perspectives of them (Gee, 1999; Lutovac & Kaasila, 2018), contributing to the development of their identity.

Every person has their own unique identity that they have developed over their lifetime, and continue to develop (Meads, 1934). This view of identity is dynamic where an individual's identity can change with new experiences and is constructed through the participation and sense of belonging in different social contexts (Gee, 1999; Kaasila, 2008; Lutovac & Kaasila, 2018; Navas, 2023; Sfard & Prusak, 2005; Wenger, 2010). Teachers who teach mathematics combine two identity aspects, their mathematical identity, and their teacher identity, to shape their mathematics-related teacher identity (Lutovac & Kaasila, 2018). The mathematical identity part relates to the relationship an individual has with mathematics, and teacher identity relates to the teaching practices and beliefs (Heyd-Metzuyanim & Shabtay, 2019). The way mathematics-related teacher identity is viewed in this study is an ongoing process constructed through participation and belonging to different groups. Telling the narratives of these experiences also contributes to the shaping of a mathematics-related teacher identity, and it is seen as the process of being and becoming (Gee, 1999; Kaasila, 2007; Sfard & Prusak; Wenger, 2010). It impacts the instruction and behavioral decisions during a mathematics lesson.

Lave and Wenger's (1991) community of practice (CoP) provides the lens for sociocultural attributes. There were three concepts of CoPs I focused on during this study: identity, trajectory, and participation. These three concepts meld together in a place where individuals can form their identity, see their trajectory (becoming), and participate within the community. There were three main CoPs my participant, Suzie, found navigating: grade team community, classroom community, and the community of the students' parents.

Methodology

The use of a narrative approach allowed me to capture the complexities of lived experiences, which bring the experiences influencing identity to the forefront (Clandinin & Connelly, 1990; Moen, 2006). Stories and poems provide the reader with a way to connect deeper with the experiences where one can feel with, not just read about (Faulkner, 2019) the events. Suzie's stories provide an understanding of how she perceives her experiences of participating and belonging in these different CoPs. These insights portray the construction of her mathematics-related teacher identity.

Suzie and I first met prior to this study when I tutored her to help pass the math portion of the general knowledge test required for her teaching certificate. During this time, we began to develop a strong, trust-based relationship. We stayed in touch as she progressed through her teacher preparation program. Later, Suzie shared the exciting news that she had been hired to teach a second-grade class, where she would be teaching all subjects. However, she also expressed some fear about teaching mathematics, as it had been a subject she struggled with in the past. I invited Suzie to participate in this research to explore how her mathematics-related teacher identity would evolve during her first year of teaching. Our prior relationship offered a unique opportunity to engage in rich, meaningful conversations. These discussions provided valuable insights into how Suzie was constructing her mathematics-related teacher's identity. This relationship is referred to as a "research friend" (Connelly & Clandinin, 1990; Kim, 2016). Further details on the impact of this relationship can be found in my dissertation (Navas, 2023).

Data and Analysis

This was a longitudinal study with multiple data collections which provided me with two perspectives, a holistic understanding of Suzie's identity and how her identity was influenced in micro-moments. During the school year, Suzie and I met online 10 times. At the beginning of each meeting, I provided Suzie "the time and space to tell her story" (Connelly & Clandinin, p. 4, 1990). I would then rephrase parts of her story to check my understanding of the events and her emotions. Additionally, I brought specific topics to discuss, often based on Suzie's recorded reflection. The structure of these meetings was flexible where Suzie was encouraged to bring up anything she wished to discuss including emotional and trying times.

Suzie also provided five recorded reflections focusing on her confidence planning and delivering a mathematics lesson. After each of our meetings, I wrote memos to capture the emotions and tone I perceived. Through this process of sharing her stories and reflecting on her experiences, Suzie continued to construct her mathematics-related teacher identity.

The meaning making process was ongoing and creative, beginning with my theoretical framework. Suzie's narratives offered me a lens to understand her evolving mathematics-related teacher identity (Gee, 1999; Kaasila, 2007). I used reflexive thematic analysis to guide my approach (Richards, 2022). Each month, I listened to, read, and reread, the transcript from our meetings, and relistened to her recorded reflections. I wrote a holistic story each month to capture the experiences and emotion of our conversations, which were then shared with Suzie. As the year progressed, I revisited each story, gaining new insights and reflections on the most meaningful experiences. This iterative process allowed me to explore how Suzie positioned herself within each CoP and how her positioning evolved over time. Through these stories, recurring ideas began to emerge, leading me to create categories that encapsulated these ideas. With the addition of new stories, I was able to add new categories that reflectively richened my understanding of their significance. I organized these categories into broader themes. Confidence and Communication were the two main themes that emerged as central to the construction of Suzie's mathematics-related teacher identity across all CoPs. I also observed that, while distinct, experiences in one CoP influenced interactions in another, highlighting their interconnectedness.

Findings

Throughout the school year, Suzie experienced moments of both struggle and success. Each CoP presented distinct challenges and interactions that shaped her planning and teaching. These CoPs influenced different facets of Suzie's overall mathematics-related teacher identity, including her mathematical identity and teacher identity.

Despite their unique contributions, the CoPs shared commonalities in how Suzie learned to become an active participant within each community. Early in the school year, Suzie expressed her appreciation for the supportiveness of her grade team, describing them as "supportive and wonderful." However, she also admitted, "I'm holding myself back because I'm too scared of saying the wrong thing." Over time, her grade team consistently encouraged her to share her ideas, which helped Suzie feel she was "being heard". By the end of the year, she reflected, "I'm not afraid to answer them, not afraid to contribute anymore… I know what I'm saying now… I've gotten a little bit more confidence speaking in the group."

Suzie’s grade team also supported her during challenging moments. Early in the year, two parents questioned her teaching decisions, which she described as a personal attack that left her questioning herself. She realized, “how important communication is,” but felt unequipped to handle the situation. Her grade team guided her on how to respond effectively and professionally. This support enabled Suzie to improve her communication with parents and establish clearer, more confident interactions.

Beyond addressing parent communication, Suzie’s grade team provided support in other areas, including classroom management and locating additional mathematics resources. At the beginning of the year, Suzie doubted her ability to teach math effectively, stating, “I feel like I’m going to do terrible.” However, as she explored new resources and adopted more intentional planning strategies, her confidence in delivering mathematics instruction grew. She remarked, “Seeing how they [the resources] teach and how they use verbiage has really helped me in my teaching… and it’s helped me be more confident in what I’m teaching.” This growing confidence was reflected in her students’ success: “Some kids who have really struggled with math get like a 100 on the quiz, and I’m just like, oh my goodness, that is just such a victory.”

Two recurring themes emerged in Suzie’s stories: the need to adapt her communication across CoPs and the development of her confidence. This confidence manifested in multiple ways, including her ability to communicate, her mathematical understanding and instructional delivery, and her classroom decision-making. To illustrate how Suzie’s past experiences influenced her current teaching, molding her confidence and communication, I present a poem from my memos based one of our conversations and her recorded reflections (Navas, 2023).

Instant dread
I can’t teach this topic
Memories of learning flood my head
Respect
Or learn.
I need to get out of my

The struggle and confusion
linger with me
I have to teach this
concept
I have to prepare
I learn
what to say

I need to feel confident in
what I teach
Otherwise,
My students won’t
Trust
but I’m better than I
thought I would be.

head
I prepare for the lesson
My confidence builds,
This concept isn't that
bad.
My students won't

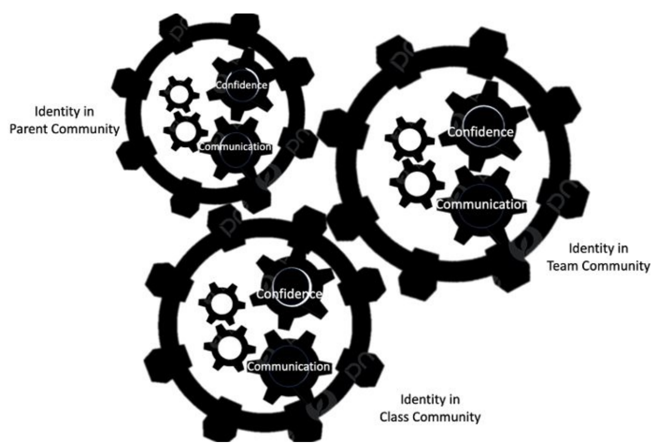
what vocabulary is
important
what should be
emphasized for success.
I may not be where I want
to be in my teaching.

Continue to grow
Continue to learn
Continue to reflect
Continue to become
An elementary
mathematics teacher.

Even though the interactions Suzie had with each CoPs were independent, the experiences she gained from one community influenced the way she participated and communicated in the other communities. The themes of communication and confidence could be seen within and between the different CoPs. For example, Suzie faced challenges communicating with the parents. Her grade team provided her with guidance on the language and tone to use while communicating with parents. When her grade team shared their knowledge, Suzie was able to communicate her expectations with confidence. This idea of the interconnectedness between and within the CoPs can be illustrated as interlinking gears that work together.

Figure 1

Gear Model of the Interconnectedness: A Mathematics-related Teacher Identity within CoPs



Conclusion

Teacher identity is a growing area of interest, as it is shaped by many factors that impact both the individual teacher and their professional practice. Exploring these factors can further deepen our understanding of how identity influences teaching and learning. This information is valuable for the mathematics education community because it sheds light on an important topic that influences all mathematics teachers at all levels. It highlights the critical role CoPs play in the construction of a teacher's mathematic-related teacher identity.

Understanding the interplay of communication and confidence within CoPs can help districts create environments where teachers can develop positive mathematics-related teacher identities such as providing environments that encourage participation. Additionally, it underscores the challenges new teachers face as they navigate communication and participation within various CoPs. This participation plays a pivotal role in shaping teacher identity and influencing the kind of educator they become. Providing teacher candidates and new teachers with opportunities to learn how to express themselves and engage meaningfully with different CoPs can provide them with knowledge and confidence to become an active participant. To assist with this knowledge, preparation and teacher induction programs could integrate activities such as composing parent e-mails. Supporting teachers' interactions with their new communities can positively impact their mathematics-related teacher identity leading to effective and equitable mathematics instruction.

References

Aragon, S. (2016). Teacher Shortages: What We Know. Teacher Shortage Series. *Education Commission of the States*.

Connelly, F. M., & Clandinin, D. J. (1990). Stories of Experience and Narrative Inquiry. *Educational Researcher*, *19*(5), 2–14. https://doi-org.ezproxy.lib.usf.edu/10.3102/0013189X019005002

Day, C., Elliot, B., & Kington, A. (2005). Reform, standards and teacher identity: Challenges of sustaining commitment. *Teaching and Teacher Education*, *21*(5), 563-577. https://doi:10.1016/j.tate.2005.03.001

Dewey, J. *Experience and Education*, Free Press, 1997. *ProQuest Ebook Central*, http://ebookcentral.proquest.com/lib/usf/detail.action?docID=4934956. Created from usf on 2021-10-22 16:19:51.

Faulkner, S. (2019). Poetic Inquiry: Craft, Method and Practice (2nd ed.). Routledge. https://doi-org.ezproxy.lib.usf.edu/10.4324/9781351044233

Gee. (1999). An introduction to discourse analysis: theory and method. Routledge.

Hodgen, J., & Askew, M. (2007). Emotion, identity and teacher learning: Becoming a primary mathematics teacher. *Oxford Review of Education*, *33*(4), https://doi.org/469-487. 0.1080/03054980701451090
Kaasila, R. (2007). Using narrative inquiry for investigating the becoming of a mathematics teacher. *Zdm*, *39*(3), 205-213.
Kim, J. (2016b). Narrative data collection methods. In J. Kim Narrative data collection methods (pp. 154-183). SAGE Publications, Inc., https://dx.doi.org/10.4135/9781071802861
Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. New York: Cambridge University Press.
Lutovac, S., & Kaasila, R. (2018). Future directions in research on mathematics-related teacher identity. *International Journal of Science and Mathematics Education, 16*(4), 759-776.
Moen, T. (2006). Reflections on the narrative research approach. *International Journal of Qualitative Methods*, *5*(4), 56-69.
Navas, K. (2023). Storying the Experiences of a First-Year Teacher's Mathematics-Related Teacher Identity with a Communities of Practice Lens (Doctoral dissertation, University of South Florida).
Richards, J. C. (2021). Coding, Categorizing, and Theming the Data: A Reflexive Search for Meaning. *Analyzing and Interpreting Qualitative Research: After the Interview*, 149.
Sfard, A., & Prusak, A. (2005). Telling identities: In search of an analytic tool for investigating learning as a culturally shaped activity. *Educational Researcher, 34*(4), 14-22.
Wenger, E. (2010). Conceptual tools for CoPs as social learning systems: Boundaries, identity, trajectories and participation. In *Social learning systems and communities of practice* (pp. 125-143). Springer, London.
Wood, M. B. (2013). Mathematical micro-identities: Moment-to-moment positioning and learning in a fourth-grade classroom. *Journal for Research in Mathematics Education, 44*(5), 775-808.