Using AI to Craft Linguistically Accessible Mathematical Tasks

Hongze Zhu, **Ri Ayat Ainul Bashirah**, **Sheida Moghtader Eslami**,
Aimaral Tabarak, **Melissa M. Soto**, **Kayla Sutcliffe**,
Hong Zhang, **& Cindy Naranjo**
University of Florida

Abstract

Creating mathematics story problems that are linguistically accessible while maintaining rigor remains a persistent challenge in diverse classrooms. This study explores how 23 elementary preservice teachers (PSTs) used ChatGPT to adapt a fourth-grade fraction multiplication problem for multilingual learners (MLLs). PSTs found ChatGPT helpful for generating adaptations quickly but noted limitations with content and visuals. The experience demonstrated how language modifications can support MLLs, positioning AI as a tool for inclusive teaching rather than a replacement for expertise.

Keywords: artificial intelligence, mathematics education, multilingual learners, task modification, elementary preservice teacher education

1 Introduction

Today's mathematics classrooms reflect the diversity of our society, with students bringing different cultural knowledge, language abilities, and life experiences. Elementary preservice teachers (PSTs) face a significant challenge: they must design mathematics tasks that are both engaging and accessible. This challenge becomes even greater in linguistically diverse classrooms (Edwards et al., 2022).

Although mathematics is often described as a "universal language," it relies on specialized vocabulary, complex grammar, and multiple modes of representation (Daroczy et al., 2015; Walkington et al., 2015). These features are especially evident in story problems. We use the term story problems instead of word problems to highlight the role of narrative thought and contextualization in mathematics, which research shows can enhance academic performance (Koedinger & Nathan, 2004; Zan, 2011). However, these same features can create barriers for multilingual learners (MLLs), even when they grasp the underlying mathematical concepts (Schleppegrell, 2007). When language masks students' true understanding, teachers might unintentionally underestimate their mathematical abilities. Research confirms that language proficiency significantly affects performance on mathematics assessments, particularly for MLLs and those from low socioeconomic backgrounds (Abedi & Lord, 2001).

For PSTs, this means developing skills not only in teaching mathematics concepts but also in adapting these concepts for students with varying English proficiency levels. This manuscript describes a lesson study in which faculty and graduate students collaborated to design an elementary mathematics methods workshop. The focus was on helping PSTs adapt mathematics story problems for MLLs across proficiency levels, with the support of GPT-4o (free version)—an AI chatbot that can generate text, answer questions, and assist with tasks based on written instructions called *prompts*. We share how generative AI (GenAI)—technology that creates new content like text or images from user requests—was used as a practical tool for modifying mathematical language to improve accessibility while maintaining rigor. Through structured engagement with ChatGPT, PSTs practiced tailoring story problems, creating sentence frames, and designing visual supports for different language stages. This work builds on research showing that adaptation of mathematical language can promote comprehension while preserving mathematical integrity (Walkington et al., 2019).

2 Design

Our team designed this workshop through a collaborative Lesson Study process (Lewis & Hurd, 2011). Together, we examined course materials, established learning goals for the PSTs, planned and taught the lesson (described below), and then reflected on both the process and PSTs' learning. Our ultimate aim was to create an experience that highlighted the linguistic challenges embedded in mathematics and demonstrated how AI tools could be leveraged to support differentiation for MLLs.

2.1 Learning Goals

We designed this lesson with several key goals in mind. First, we wanted PSTs to recognize how students' language proficiency levels influence their ability to engage in mathematics learning and to consider strategies for supporting MLLs in participating in rigorous mathematical tasks at each level of proficiency. Language proficiency can significantly shape how students comprehend story problems and how they respond to questions from teachers and peers. Table 1 outlines these proficiency levels (Krashen, 1984), describes students' typical attributes at each stage, and provides examples of academic supports PSTs can use to scaffold MLLs' learning across language skills (WIDA, 2025).

Table 1. *Ideas for Language Supports Across Literacy Skills*

Level of Proficiency	Attributes	Reading	Writing	Listening	Speaking
Pre-Production	Focus is on absorbing English but not speaking independently	Ask students to match symbols with visuals.	Ask students to use drawings to communicate meaning.	Ask students to point to objects and/or symbols.	Ask students yes/no questions, have them repeat key terms.
Early Production	Answer basic questions with one- or two-word responses or short phrases	Ask students to match key terms with visuals.	Ask students to fill in blanks and/or label objects.	Ask students to follow simple, one- to two-word directions.	Ask students questions that they can answer with one- or two-word responses.
Speech Emergence	Can produce phrases and short sentences	Ask students to read simple story problems with visual supports.	Ask students to write simple sentences to explain their thinking.	Students can follow multi-step directions.	Ask students to use sentence frames to describe basic steps.
Intermediate Fluency	Have good social communication skills, need support with academic language	Ask students to read multi-step problems with some visual supports.	Ask students to write using more precise vocabulary and explain their thinking with minimal support.	Students can understand more complex oral directions.	Ask students to use sentence frames to produce more complex verbal explanations.

Second, we aimed to help PSTs develop practical strategies for modifying story problems for different proficiency levels. Because MLLs vary in their stages of second language proficiency, teachers can use this knowledge to differentiate instruction for MLLs in reading, writing, listening, and speaking for academic purposes. Third, we wanted PSTs to gain experience using AI tools thoughtfully in their planning process. Last, we wanted PSTs to engage with a mathematical story problem involving the multiplication of whole numbers and fractions. This allowed them to anticipate both the solution strategies students might use and the language-related challenges students could encounter when working through the task.

Many teacher candidates have limited experience with MLLs (Lucas et al., 2014; Sugimoto et al., 2017), so we developed activities that would build their awareness of language demands in mathematics tasks. By practicing with ChatGPT to modify story problems, PSTs gained firsthand experience with how small language adjustments can make mathematical content more accessible without reducing cognitive demand. Our ultimate goal extends beyond this single lesson. We hope to nurture future teachers who create mathematics experiences that value students' linguistic and cultural backgrounds while maintaining high expectations for mathematical thinking, aligning with the National Council of Teachers of Mathematics Intersection of Culture and Mathematics statement (NCTM, 2024). As PSTs become more aware of language features in mathematics problems, they can better support all students in developing language and mathematical understanding simultaneously.

2.2 Lesson Plan

Our workshop spanned two sessions in two days. In the first session (50 minutes), PSTs worked in small groups to read the following story problem, "At the big table, 12 kids shared 8 bottles of juice so that everyone got the same amount of juice. There are 3 kids sitting at the little table. How much juice should we put on the little table if we want each kid at the little table to have as much juice as each kid at the big table?" They then read four ChatGPT-modified versions of the story problem that contained additional instructional supports and sort them by language proficiency level (Table 2).

Table 2. *Story Problem Modifications Across Language Proficiency Levels*

Proficiency Level	Modified Story Problem	Additional Instructional Supports
Pre-Production (Spanish example)	En la mesa grande, 12 niños compartieron 8 botellas de jugo, de manera que todos recibieron la misma cantidad de jugo. Hay 3 niños sentados en la mesa pequeña. ¿Cuánto jugo debemos poner en la mesa pequeña si queremos que cada niño de esa mesa tenga la misma cantidad que cada niño de la mesa grande?	• Translation in L1[1] • Pictures of Key Vocabulary • Highlighted Vocabulary • Spanish audio recording (pre-L1) • Written Spanish translation (L1)
Early Production (Spanish example)	En la table grande, 12 niños compartieron 8 botellas de juice de manera que cada uno recibió la misma cantidad de juice. Hay 3 niños sentados en la table pequeña. ¿Cuánto juice debemos poner en la table pequeña si queremos que cada niño en la table pequeña reciba la misma cantidad de juice que cada niño en la table grande?	• Translation in L1 • Pictures of Key Vocabulary • English Translation with Highlighting
Speech Emergence	At the big table, 12 kids shared 8 bottles of juice equally. Now, 3 kids are at the little table. They want the same amount of juice. How many bottles of juice do we need at the little table?	• Pictures of Key Vocabulary • Shorter Sentences • Assessing Questions • Sentence Frames
Intermediate Fluency	At the big table, 12 kids shared 8 bottles of juice equally. Now, 3 kids are at the little table. They want the same amount of juice. How many bottles of juice do we need at the little table?	• Word Bank with Pictures

Although ChatGPT provided an initial set of story problems across proficiency levels, the problems were not used without modification. The Spanish translations were especially helpful, but additional revisions were required to ensure appropriateness for students at varying stages of English language development. For each problem, they discussed which stage of English development (e.g., pre-production, early production, intermediate) the problem was most appropriate for and identified linguistic features that informed their decision.

[1]L1 = First Language

The second session (two hours) began with mathematical problem-solving. We asked PSTs to solve our focus problem using at least two different strategies appropriate for fourth graders. This activity helped them anticipate student thinking and develop a deeper understanding of fraction multiplication.

Throughout both sessions, we emphasized that modifications should maintain mathematical rigor while making language more accessible. The goal is not to simplify the mathematics, but rather to reduce unnecessary linguistic complexity that might interfere with students' ability to demonstrate their mathematical understanding.

2.3 Math Task Selection

The foundational mathematics task we selected for the second session was a story problem from our state's fourth grade mathematical standards' "Instructional Guide for Mathematics" document (Florida Department of Education, 2023): "The theater class needs $\frac{3}{4}$ foot of ribbon for each costume. They need to make 8 costumes. How many feet of ribbon are required in total?" This problem offered several valuable features for our workshop. First, it addressed the mathematical goal of multiplying fractions by whole numbers, an important fourth-grade concept that many students (and teachers) find challenging (Van Steenbrugge et al., 2014). The problem naturally invited multiple solution strategies, from concrete approaches like drawing diagrams to more abstract methods such as complex doubling strategies (Baek, 2006).

Prior to solving the problem, PSTs identified key information, discussed what the numbers represented, and considered how fourth graders might approach the problem connecting to the state standards for fraction multiplication. The instructor then prompted: "Anticipate at least two different strategies 4th graders would use to solve this problem. And think about what might be challenging for MLLs when engaging in this problem?"

Figures 1 and 2 show different solution approaches from PSTs. Figure 1 demonstrates a traditional multiplication approach paired with a repeated addition strategy, showing how students might add $\frac{3}{4}$ eight times to reach 6 feet. Figure 2 illustrates a contrasting approach in which the PST used visual aids. The student drew representations of the problem and worked with $\frac{1}{4}$ units to determine how much ribbon would not be used, ultimately arriving at the final answer. These varied approaches highlight the different ways fourth graders might conceptualize the multiplication of fractions by whole numbers, while also demonstrating visual models that could support MLLs' understanding of the problem.

Figure 1

PST's handwritten solution to the theater ribbon problem showing multiplication strategy ($\frac{3}{4} \times 8$) and repeated addition of eight fractions ($\frac{3}{4} + \frac{3}{4} + \ldots$), with fraction bars representing ribbon

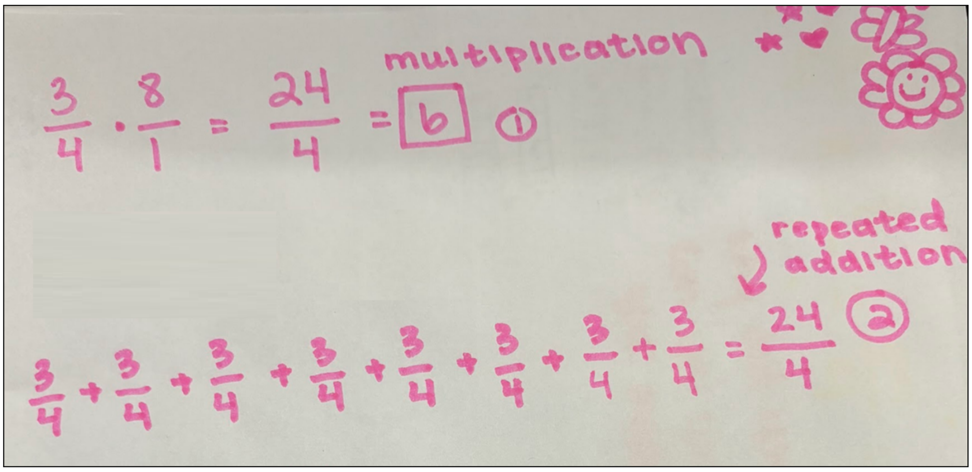
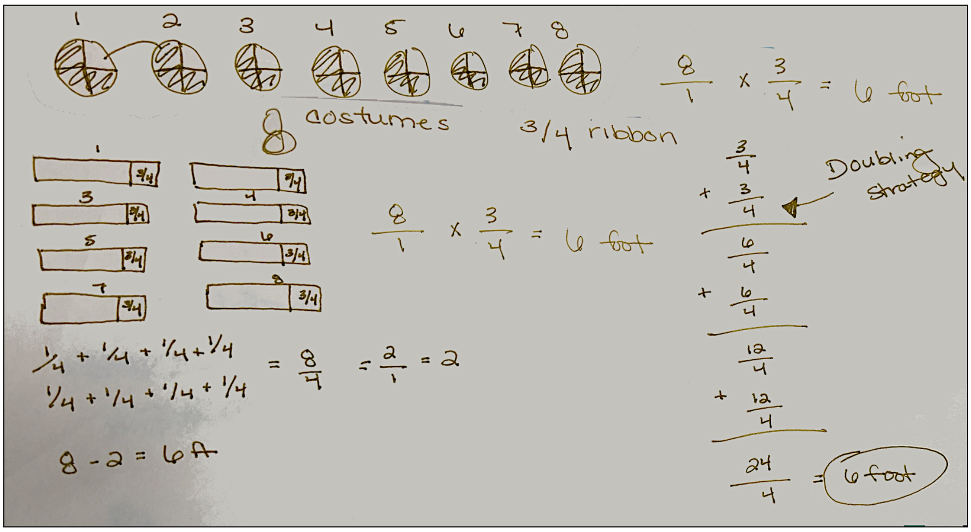


Figure 2
PST's handwritten solution showing fraction multiplication using circle models and a doubling strategy

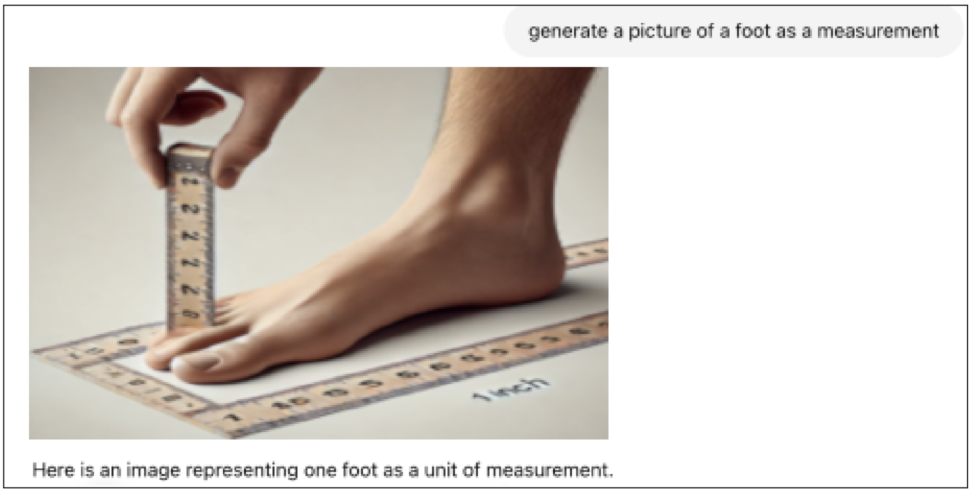


This problem-solving experience led to discussions about both mathematical and linguistic challenges that MLLs might face. While mathematically straightforward, the problem contains several linguistic features that might challenge MLLs. These include:

- Specialized vocabulary (theater, costume, ribbon, require)
- The irregular plural transformation from "foot" to "feet"
- Technical terms (in total, each)
- Passive voice structure
- Implied mathematical operations

The problem also introduced cultural elements that might be unfamiliar to some students. Theater productions and costume-making might be common experiences for some students but completely unfamiliar to others. Similarly, measurements in feet might confuse students from countries that use the metric system and even some native English speakers who may confuse the homonym for a unit of measurement or a body part. This term also confused ChatGPT when PSTs asked it to create a picture of a foot of measurement (see Figure 3).

Figure 3
ChatGPT's Misinterpretation of 'Foot' as a Body Part Rather Than a Unit of Measurement



By analyzing these language features with PSTs, they recognized potential barriers in seemingly simple word problems and considered modifications that might make the mathematics more accessible to students at different language proficiency levels.

3 AI Integration Process

After solving the story problem, but before introducing ChatGPT, we surveyed the group about their familiarity with AI tools to gauge how much support they might need. Results showed that most PSTs had at least some experience: 12 reported using ChatGPT or other GenAI tools "sometimes" for schoolwork, 7 used them "often," and only 3 had little to no prior experience. In addition, 13 PSTs described themselves as comfortable with AI, suggesting they were ready to explore its role in mathematics instruction. When asked how they might use ChatGPT as future teachers, their responses clustered around several themes: creating instructional materials (lesson plans, worksheets, assessments), generating engaging activities, modifying tasks for multilingual learners, and supporting their own content understanding. Notably, several PSTs specifically mentioned using AI to help with accommodations for MLLs, reflecting an emerging awareness of its potential to support linguistically diverse classrooms.

3.1 Guided Exploration with ChatGPT

With this foundation established, we moved into the collaborative exploration of ChatGPT's capabilities. Working in small groups, PSTs responded to the instructor's directive: "Focus on adjusting the story problem to address one of the highest levels of language proficiency (Intermediate Fluency). Develop instructional supports and assessing and advancing questions that MLLs can respond to at this proficiency level."

PSTs began by experimenting with the AI tool to modify the theater ribbon problem for students at the intermediate fluency level. This process required PSTs to balance mathematical rigor with linguistic accessibility. Groups spent fifteen minutes having conversations with ChatGPT, learning how to craft effective prompts and critically evaluating the generated content. They focused on creating appropriate instructional supports such as visual representations, sentence frames, simplifying sentence structures, and vocabulary supports that aligned with students' language proficiency. We intentionally began with the intermediate fluency level because tasks and instructional supports at this stage remain appropriate for most native English speakers while also becoming more accessible for MLLs. This practice helped PSTs see how thoughtful modifications could create inclusive mathematics experiences for diverse classrooms.

3.2 Independent Practice with ChatGPT

After completing the small-group exploration, PSTs then moved into independent practice, continuing to use ChatGPT to modify and adapt tasks on their own. This final phase emphasized independent application, challenging PSTs to adapt their instructional supports for students at the speech emergence level. Figures 4, 5, and 6 illustrate how one group of PSTs engaged with ChatGPT during this process. In Figure 4, PSTs requested specific modifications for English learners at the speech emergence level, with ChatGPT suggesting simplified language, visual supports, and guiding questions appropriate for this proficiency level. Figure 5 shows a visual aid generated by ChatGPT—a theater scene with costumes and ribbon measurements—designed to make mathematical quantities more concrete for students with limited English proficiency. Despite its attempt to support understanding, the image demonstrates limitations in AI-generated visuals, with measurements that were difficult to read and inconsistent representations of fractions. More specifically, for example, the story problem called for making 8 costumes, yet the ChatGPT-generated image depicted 12 costumes—an inconsistency that did not reflect the task and risked confusing learners.

In Figure 6, PSTs refined their approach by explicitly asking for visual aids to support vocabulary comprehension, demonstrating their developing understanding of how to scaffold language through multiple modalities. ChatGPT responded with specific suggestions for representing key terms like "costume" and "ribbon" visually.

Figure 4

PST-ChatGPT Conversation about Modifying Tasks for Speech Emergence Level

PST said:
How would you modify the following word problem for an English learner at the speech emergence level: The theater class has to make costumes that require ¾ foot of ribbon. They need to make 8 costumes. How many feet of ribbon do they need total?
AI said:
To modify this problem for an English learner at the speech emergence level, simplify the language and include visuals or context to support understanding. Here's an example of how you might rewrite it:
Simplified Problem:
The theater class is making costumes. Each costume needs ¾ **foot of ribbon**. They will make **8 costumes**.
How much ribbon do they need in total?
Tips:

- Provide a picture of ribbon or costumes to give context.
- Write "¾ foot" as both text and a fraction.
- Use a number line or visual fractions to support understanding if possible.
- Allow them to use hands-on tools like rulers or strips of paper representing ¾ foot of ribbon.

For example:

- **Picture:** Show 8 small costumes and a piece of ribbon labeled "¾ foot" next to each one.
- Include guiding questions like:
  - "What does ¾ foot mean?"
  - "If 1 costume needs ¾ foot, how many ribbons do we need for 8 costumes?"

Figure 5

ChatGPT-Generated Visual Support for the Theater Ribbon Problem

Figure 6

PST-ChatGPT Conversation about Developing Visual Supports

PST said:
What pictures would you include to help the student understand unfamiliar vocabulary in the problem
AI said:
To help an English learner understand unfamiliar vocabulary, include **simple and clear pictures** that visually represent key terms. Here's what you could include:

- **Costume**
  - A picture of a basic costume, like a dress, cloak, or theater outfit, to clarify what the word "costume" means in this context.
- **Ribbon**
  - A close-up image of a ribbon, possibly with a ruler next to it to show the concept of measurement.
- **¾ foot**
  - A visual representation of ¾ as part of a whole. For example:
    - A ruler divided into 4 parts, with 3 parts shaded or labeled.
    - A strip of paper labeled "1 foot," with ¾ of it highlighted or cut out.
- **Theater Class**
  - An image of students working together on costumes or performing on stage, connecting the "class" with the idea of making costumes.
- **Multiplication (Optional)**
  - A visual of repeated addition or groups (e.g., 8 ribbons, each labeled ¾ foot) to introduce multiplication as a faster method of adding the same number repeatedly.

We concluded the lesson with a whole-class discussion and self-assessment, asking PSTs to compare the challenges they faced when developing tasks for different proficiency levels. This reflection helped identify areas where PSTs needed additional support and built awareness of how AI tools can both help and hinder the differentiation process.

Throughout all phases, we emphasized critical reflection on AI's role in teaching. We wanted PSTs to recognize that while ChatGPT can efficiently generate ideas and language modifications, human judgment remains essential. PSTs evaluated whether AI-generated content maintains mathematical rigor, addressed the intended learning goals, and provided appropriate support for the specific language needs of their students. This structured approach helped PSTs develop both technical skills with AI tools and the critical thinking needed to use them effectively in planning mathematics instruction for linguistically diverse classrooms.

3.3 Methodology and Workflow

We selected our foundational task from our state's fourth-grade Instructional Guide for Mathematics, choosing a problem that addressed fraction multiplication while offering multiple entry points for student thinking. The theater ribbon problem asked students to calculate how much ribbon would be needed to make eight costumes when each costume required three-fourths of a foot. We chose this particular problem because it involved mathematics that both students and teachers find challenging, yet it was accessible enough to allow for various solution strategies. The problem also contained language features that often create barriers for MLLs, making it an ideal candidate for exploring how ChatGPT might support task modification.

Our process for working with ChatGPT followed a clearly defined sequence. We began by having PSTs solve the original problem themselves and analyze its linguistic features prior to introducing the AI tool. This work helped them develop a clear understanding of the mathematics and potential language barriers, which proved essential when later evaluating ChatGPT's suggestions. We then moved into guided exploration, asking small groups to modify the task for students at the intermediate fluency level. We chose to start with this higher proficiency level intentionally, as modifications at this stage tend to benefit all students while still enhancing accessibility for MLLs. Groups spent about fifteen to twenty minutes conversing with ChatGPT, learning how to write effective prompts and critically evaluate the generated content. After sharing their work, the groups transitioned to independent practice, adapting the problem for students at the speech emergence level.

Throughout this process, we emphasized that working with ChatGPT is an iterative process rather than an immediate solution. PSTs quickly learned that they needed to refine their prompts, specify their goals clearly, and remain careful about reviewing the tool's output for accuracy. We asked them to save their ChatGPT conversations so we could later examine how they communicated with the tool and what modifications emerged. In the following section, we will describe PSTs' perceptions in using ChatGPT to request specific supports, such as sentence frames, visual aids, and vocabulary scaffolds as well as how they approached the use of GenAI as a tool. Specifically, on how to think critically about what makes mathematics problems comprehensible for students at different language proficiency levels.

4 Preservice Teachers' Voice

At the end of the second session of the lesson, PSTs completed a post-lesson reflection on Google Forms, which included the following questions:

1. How did using ChatGPT support you in creating comprehensible math story problems for the various levels of MLLs' proficiencies?
2. What benefits and challenges did you experience in the process of making these math story problems?
3. What insights did you gain about supporting multilingual learners' needs?

The researchers used inductive thematic analysis (Braun & Clarke, 2006) to code the data. The reflections were open-coded by question to identify recurring ideas, language, and insights. Codes were then clustered into broader categories to develop emergent themes.

Reflections after the workshop revealed how PSTs experienced using AI to modify story problems for MLLs. Many PSTs highlighted ChatGPT's ability to simplify language in ways they might not have considered independently. One PST noted how ChatGPT transformed phrases like "required in total" into more accessible alternatives like "altogether." The generation of scaffolding resources emerged as a significant benefit in PSTs' reflections. ChatGPT helped create sentence frames that guided students through problem-solving processes and developed step-by-step problem breakdowns. One PST wrote that it was "great for creating sentence frames and ideas for accessing questions to ask," while another valued how it "helped organize my thinking in a step-by-step process." PSTs also recognized ChatGPT's utility in developing word banks with pictures, creating guiding questions that followed a sensible progression through problems, and providing translation options for key mathematical terms, applications that resonate with AI's emerging role in personalized and adaptive learning (Supriyadi & Kuncoro, 2023).

Despite these advantages, PSTs consistently identified limitations in ChatGPT's capabilities. Image generation proved particularly problematic, with one PST noting that "the imagery was a huge disadvantage" and another stating that they would "stick to Google from now on for image production." Several commented that while ChatGPT could suggest what visuals might be useful, the actual images it generated were confusing or inappropriate for classroom use. PSTs also emphasized the iterative nature of working with AI tools. Many described needing to "keep correcting ChatGPT" and being "very specific" with prompts to achieve desired results. One PST observed that "it took several times to create a final product," while another explained, "we had to keep asking it to continue breaking the question down." This suggests that effective use of AI requires pedagogical knowledge about how to communicate instructional goals clearly and adapt when technology does not immediately produce desired results (Antonenko & Abramowitz, 2023). Educators need both discipline-specific expertise and general educational knowledge to successfully integrate AI tools into their practice.

Perhaps most significantly, many PSTs indicated that the experience enhanced their understanding of supporting MLLs. "I gained insight on how important word choice and visuals are in supporting MLLs' needs," wrote one PST. Another reflected, "I learned that there are different levels of MLL and their abilities," while a third noted learning "how to better adapt my word problems for the different stages of learning including those who do not have [English] literacy yet." These insights highly align with Bailey & Heritage's (2014) "language learning progressions," which recognize varying language proficiency levels and diverse developmental pathways, enabling more appropriate expectations and supporting teachers in tailoring instruction to students' current abilities.

Overall, PSTs came to view AI as a complementary tool rather than a replacement for teacher expertise, aligning with research showing that while teachers recognize AI's potential to enhance instruction, they emphasize the irreplaceable role of human educators in addressing ethical concerns, adapting to student needs, and fostering critical thinking (Antonenko & Abramowitz, 2023). As one PST summarized, "I appreciated learning about a tool that can brainstorm ideas for me and I can expand on using my own knowledge." Another expressed similar sentiment: "It helped me think of ways to modify the problems so that each student can be supported." These reflections suggest that while AI can serve as a valuable starting point for task modification, teachers must critically evaluate and adapt AI-generated content to effectively support their MLLs' specific needs.

Reflections after the workshop revealed how PSTs experienced using AI to modify story problems for MLLs. Many PSTs highlighted ChatGPT's ability to simplify language in ways they might not have considered independently. One PST noted how ChatGPT transformed phrases like "required in total" into more accessible alternatives like "altogether." The generation of scaffolding resources emerged as a significant benefit in PSTs' reflections. ChatGPT helped create sentence frames that guided students through problem-solving processes and developed step-by-step problem breakdowns. One PST wrote that it was "great for creating sentence frames and ideas for accessing questions to ask," while another valued how it "helped organize my thinking in a step-by-step process." PSTs also recognized ChatGPT's utility in developing word banks with pictures, creating guiding questions that followed a sensible progression through problems, and providing translation options for key mathematical terms, applications that resonate with AI's emerging role in personalized and adaptive learning (Supriyadi & Kuncoro, 2023). Despite these advantages, PSTs consistently identified limitations in ChatGPT's capabilities. Image generation proved particularly problematic, with one PST noting that "the imagery was a huge disadvantage" and another stating that they would "stick to Google from now on for image production." Several commented that while ChatGPT could suggest what visuals might be useful, the actual images it generated were confusing or inappropriate for classroom use. PSTs also emphasized the iterative nature of working with AI tools. Many described needing to "keep correcting ChatGPT" and being "very specific" with prompts to achieve desired results. One PST observed that "it took several times to create a final product," while another explained, "we had to keep asking it to continue breaking the question down." This suggests that effective use of AI requires pedagogical knowledge about how to communicate instructional goals clearly and adapt when technology does not immediately produce desired results (Antonenko & Abramowitz, 2023). Educators need both discipline-specific expertise and general educational knowledge to successfully integrate AI tools into their practice. Perhaps most significantly, many PSTs indicated that the experience enhanced their understanding of supporting MLLs. "I gained insight on how important word choice and visuals are in supporting MLLs' needs," wrote one PST. Another reflected, "I learned that there are different levels of MLL and their abilities," while a third noted learning "how to better adapt my word problems for the different stages of learning including those who do not have [English] literacy yet." These insights highly align with Bailey & Heritage's (2014) "language learning progressions," which recognize varying language proficiency levels and diverse developmental pathways, enabling more appropriate expectations and supporting teachers in tailoring instruction to students' current abilities. Overall, PSTs came to view AI as a complementary tool rather than a replacement for teacher expertise, aligning with research showing that while teachers recognize AI's potential to enhance instruction, they emphasize the irreplaceable role of human educators in addressing ethical concerns, adapting to student needs, and fostering critical thinking (Antonenko & Abramowitz, 2023). As one PST summarized, "I appreciated learning about a tool that can brainstorm ideas for me and I can expand

on using my own knowledge." Another expressed similar sentiment: "It helped me think of ways to modify the problems so that each student can be supported." These reflections suggest that while AI can serve as a valuable starting point for task modification, teachers must critically evaluate and adapt AI-generated content to effectively support their MLLs' specific needs.

5 Discussion

New technologies typically follow three processes as they become integrated into educational practice: resistance, adaptation, and reconstruction (Geels, 2002). We have seen this pattern with calculators, which initially faced resistance before becoming accepted as valuable instructional resources. AI tools present a similar but more powerful and complex transition, as they offer greater capabilities while raising more significant concerns about potentially diminishing learning opportunities. By exploring AI through multiple perspectives and carefully considering both its potential and its limitations, we can move toward thoughtful adaptation of these tools in mathematics education. Our workshop on AI-assisted modifications for mathematical story problems offers valuable insights for teacher preparation and mathematics instruction, addressing a pressing need: preparing teachers to make mathematics accessible to diverse learners while maintaining cognitive demand.

5.1 Teacher Preparation Programs

As AI increasingly permeates our daily lives, teacher education programs must prepare future educators to thoughtfully integrate these tools into their planning process, empowering them to decide how and when to use these tools in their educational practice (Antonenko & Abramowitz, 2023). Our workshop revealed that PSTs need specific skills to effectively use AI for linguistically responsive mathematics teaching. First, PSTs need guidance in evaluating and adapting AI-generated content (Allen et al., 2024). As our participants discovered, ChatGPT can quickly produce modifications, but these modifications often require refinement. Teacher educators should help PSTs develop critical assessment skills to determine when AI suggestions enhance or potentially detract from student learning (Allen et al., 2024). This includes recognizing when language simplifications maintain mathematical integrity and when they inadvertently remove important conceptual elements.

Second, PSTs need experience maintaining mathematical rigor while enhancing linguistic accessibility (Roberts & Olarte, 2023). Our workshop deliberately focused on modifying language features without reducing cognitive demand. This balance is essential but challenging to achieve. When working with our theater ribbon problem, we asked PSTs to solve the problem first because this helped them understand the level of cognitive demand. Teacher preparation programs should provide structured opportunities for PSTs to practice this delicate balance.

Third, PSTs benefit from learning to develop targeted supports for different language proficiency levels. By working with the same problem across multiple proficiency levels, our PSTs gained insight into how supports must evolve as students develop language skills. Teacher preparation programs should provide interdisciplinary courses to help PSTs understand language acquisition progressions and connect these to mathematics instruction (Bailey & Heritage, 2014). This knowledge helps teachers avoid both oversimplification which reduces mathematical learning and overly complex language that creates unnecessary barriers to understanding (Healy & Powell, 2012).

5.2 Mathematics Instruction

For practicing teachers, AI tools offer promising avenues for enhancing mathematics instruction in linguistically diverse classrooms (NCTM, 2024). Our work suggests two primary applications worth exploring. First, AI can help teachers efficiently generate differentiated mathematical tasks (Supriyadi & Kuncoro, 2023). Many novice teachers recognize the need for differentiation but struggle with the time demands (Fantilli & McDougall, 2009). As our workshop demonstrated, AI tools can quickly generate modifications that address specific language needs while maintaining mathematical goals. This efficiency could make differentiation more feasible in busy classroom settings. However, teachers must recognize that AI-generated problems require careful review to ensure they accurately present mathematical concepts and avoid introducing confusion (Sawyer, 2024).

Second, AI can assist in creating appropriate scaffolds and supports (Sawyer, 2024). Our PSTs found ChatGPT particularly helpful in generating sentence frames, word banks, and sequenced questioning strategies tailored to different proficiency levels. These scaffolds can support student engagement with rigorous mathematical content while developing both mathematical and language understanding. Rather than viewing language and mathematics learning as separate endeavors, this approach recognizes their interconnectedness and supports each other.

6 Implications

6.1 Tips for PSTs/Teachers/Mathematics Teacher Educators

We conclude with suggestions for making mathematical tasks more comprehensible for MLLs. First, these groups should focus on getting to know their students, anticipate solution strategies, and considering varying language proficiencies and individual differences so they can provide targeted support and critically evaluate AI outputs. Second, PSTs, in-service teachers, and MTEs should spend time exploring the GenAI tools, learning benefits and limitations about the different platforms to make the most of it and to reduce the novelty of the technology (Chan & Lee, 2023). We found that in our workshop, some PSTs were off topic asking ChatGPT to generate pictures not related to the task. By providing time to explore the tool, it can reduce the novelty effect. Third, GenAI can be helpful in generating additional instructional supports such as sentence frames, question prompts, pictures, and ideas. However, PSTs, in-service teachers, and MTEs should be specific and realize it is an iterative process of refining their prompts. Also, at times, it is easier to search or ask GenAI to search for existing visuals rather than asking it to create them. Additionally, these instructional supports are beneficial to all learners and not just MLLs. Lastly, collaboration with diverse professionals is essential for creating effective lessons, as advocated by NCTM (2025). In our workshop, we included MTEs and ESOL TEs in our Lesson Study team to draw upon their expertise in supporting MLLs.

6.2 Limitations & Future Directions

Our study explored how PSTs used ChatGPT to modify mathematical story problems for multilingual learners (MLLs), with our analysis focusing on both the process of working with AI and PSTs' reflections on their experiences. While we did not systematically evaluate the instructional quality or linguistic accessibility of the AI-generated tasks in this manuscript, this remains an important direction for future work.

Moving forward, teachers and teacher educators would benefit from clear frameworks for evaluating AI-generated modifications. These should help determine whether outputs preserve mathematical accuracy, reduce language complexity appropriately, and remain culturally relevant. Classroom-based studies are also needed to examine how MLLs themselves engage with AI-modified tasks and whether such adaptations meaningfully support both mathematical understanding and language development.

Another important next step is to provide PSTs with models of what high-quality AI outputs should look like—for example, through checklists or example prompts and responses. Such tools would help PSTs critically evaluate AI suggestions and make informed instructional decisions. More broadly, as AI tools continue to evolve, educators need sustained support to use them effectively and ethically. This includes developing shared guidelines for AI use in educational planning and ensuring equitable access to these technologies across diverse school settings. Ultimately, the goal is not to replace teachers' expertise but to strengthen it—giving educators new tools to make mathematics accessible, rigorous, and meaningful for every learner.

References

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*(3), 219–234. https://doi.org/10.1207/S15324818AME1403_2

Allen, T., Villaflor-Wilson, R., Muljana, P., & Romero-Hall, E. (2024). AI-generated content: Guidelines, higher-order thinking skills, and copyrights. *Educational Technology Journal*, *4*(1), 1–5.

Antonenko, P., & Abramowitz, B. (2023). In-service teachers' (mis)conceptions of artificial intelligence in K-12 science education. *Journal of Research on Technology in Education*, *55*(1), 64–78. https://doi.org/10.1080/15391523.2022.2119450

Baek, J. M. (2006). Research, reflection, practice: Children's mathematical understanding and invented strategies for multidigit multiplication. *Teaching Children Mathematics*, *12*(5), 242–247.

Bailey, A. L., & Heritage, M. (2014). The role of language learning progressions in improved instruction and assessment of English language learners. *TESOL Quarterly*, *48*(3), 480–506. https://doi.org/10.1002/tesq.176

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chan, C. K. Y., & Lee, K. K. (2023). The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and millennial generation teachers? *Smart Learning Environments*, *10*(1), 60. https://doi.org/10.1186/s40561-023-00269-3

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H. C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology*, *06*. https://doi.org/10.3389/fpsyg.2015.00348

Edwards, M. T., Yang, Z., & Zeng, J. (2022). Fostering mathematical inquiry with language independent board games. *Ohio Journal of School Mathematics*, *91*, 17–28.

Fantilli, R. D., & McDougall, D. E. (2009). A study of novice teachers: Challenges and supports in the first years. *Teaching and Teacher Education*, *25*(6), 814–825. https://doi.org/10.1016/j.tate.2009.02.021

Florida Department of Education. (2023). *Instructional Guide for Mathematics*. Florida Department of Education.

Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: A multi-level perspective and a case-study. *Research Policy*, *31*(8–9), 1257–1274. https://doi.org/10.1016/S0048-7333(02)00062-8

Healy, L., & Powell, A. B. (2012). Understanding and overcoming "disadvantage" in learning mathematics. In M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third International Handbook of Mathematics Education* (pp. 69–100). Springer New York. https://doi.org/10.1007/978-1-4614-4684-2_3

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *The Journal of the Learning Sciences*, *13*(2), 129–164. https://doi.org/10.1207/s15327809jls1302_1

Krashen, S. D. (1984). *Principles and practice in second language acquisition* (Reprinted). Pergamon Press.

Lewis, C. C., & Hurd, J. (2011). *Lesson study step by step: How teacher learning communities improve instruction*. Heinemann.

Lucas, T., Villegas, A. M., & Martin, A. D. (2014). Teachers' beliefs about English language learners. In *International Handbook of Research on Teachers' Beliefs* (pp. 453–474). Routledge.

National Council of Teachers of Mathematics. (2024). *The intersection of culture and mathematics: A position of the National Council of Teachers of Mathematics.* https://www.nctm.org/Standards-and-Positions/Position-Statements/The-Intersection-of-Culture-and-Mathematics/

National Council of Teachers of Mathematics. (2025). *Changing the professional culture of teaching mathematics: A position of the National Council of Teachers of Mathematics.* https://www.nctm.org/Standards-and-Positions/Position-Statements/Changing-the-Professional-Culture-of-Teaching-Mathematics/

Roberts, S. A., & Olarte, T. R. (2023). Enacting multilingual learner core practices: A PST's approximations of practice of mathematics language routines. *Journal of Mathematics Teacher Education.* https://doi.org/10.1007/s10857-023-09600-8

Sawyer, A. G. (2024). Artificial intelligence chatbot as a mathematics curriculum developer: Discovering preservice teachers' overconfidence in ChatGPT. *International Journal on Responsibility*, *7*(1). https://doi.org/10.62365/2576-0955.1106

Schleppegrell, M. J. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading & Writing Quarterly*, *23*(2), 139–159. https://doi.org/10.1080/10573560601158461

Sugimoto, A. T., Carter, K., & Stoehr, K. J. (2017). Teaching "in their best interest": Preservice teachers' narratives regarding English Learners. *Teaching and Teacher Education*, *67*, 179–188. https://doi.org/10.1016/j.tate.2017.06.010

Supriyadi, E., & Kuncoro, K. S. (2023). Exploring the future of mathematics teaching: Insight with ChatGPT. *Union: Jurnal Ilmiah Pendidikan Matematika*, *11*(2), 305–316. https://doi.org/10.30738/union.v11i2.14898

Van Steenbrugge, H., Lesage, E., Valcke, M., & Desoete, A. (2014). Preservice elementary school teachers' knowledge of fractions: A mirror of students' knowledge? *Journal of Curriculum Studies*, *46*(1), 138–161. https://doi.org/10.1080/00220272.2013.839003

Walkington, C., Clinton, V., & Sparks, A. (2019). The effect of language modification of mathematics story problems on problem-solving in online homework. *Instructional Science*, *47*(5), 499–529. https://doi.org/10.1007/s11251-019-09481-6

Walkington, C., Clinton, V., Ritter, S. N., & Nathan, M. J. (2015). How readability and topic incidence relate to performance on mathematics story problems in computer-based curricula. *Journal of Educational Psychology*, *107*(4), 1051–1074. https://doi.org/10.1037/edu0000036

WIDA. (2025, August 4). *Can do descriptors.* WIDA: University of Wisconsin-Madison. https://wida.wisc.edu/teach/can-do/descriptors

Zan, R. (2011). The crucial role of narrative thought in understanding story problems. *Current State of Research on Mathematical Beliefs XVI*, 287–305.

Hongze Zhu, **Ri Ayat Ainul Bashirah**, **Sheida Moghtader Eslami**, **Aimaral Tabarak**, **Melissa M. Soto**, **Kayla Sutcliffe**, **Hong Zhang**, and **Cindy Naranjo** are colleagues in the School of Teaching and Learning at the University of Florida, where they collaborate on research related to mathematics education, multilingual learners, and the use of AI in teacher preparation.

A Quick-Start Guide: Using GenAI to Modify Math Tasks for Multilingual Learners (MLLs)

1. **Get to Know Your Students.** Start by understanding your students' language proficiency, background knowledge, and problem-solving strengths. Use this information to guide how you adapt mathematical language, visuals, and instructions.

2. **Learn the AI Tool First.** Spend time exploring GenAI tools (like ChatGPT) before using them for lesson planning. Experiment with prompts, learn what the tool can and cannot do, and reduce the "novelty effect" so you can use it purposefully.

3. **Use GenAI for Targeted Supports.** Ask GenAI to generate specific supports, such as sentence frames, vocabulary lists, question prompts, and visuals, to make math tasks more accessible. Review all outputs carefully for clarity and accuracy.

4. **Refine and Evaluate Outputs.** Treat GenAI as a collaborator, not a substitute. Revise and refine its suggestions iteratively. Check that each output is mathematically correct, linguistically accessible, and culturally relevant for your learners.

5. **Use AI Strategically.** At times, it may be faster or more effective to search for an existing image rather than asking AI to create one. Use your professional judgment to decide when AI adds real value.

6. **Collaborate with Colleagues.** Work with other teachers, ESOL specialists, and math educators to co-design and evaluate AI-supported lessons. Collaboration helps ensure that all adaptations are appropriate for multilingual learners and aligned with standards.

7. **Focus on Equity and Accessibility.** Remember that GenAI is a tool to make math instruction more inclusive. Use it to help all students, especially MLLs, access rigorous content, engage meaningfully, and develop both mathematical and language proficiency.

Acknowledgments

We extend our sincere gratitude to Dr. Kristen Apraiz, who participated in multiple lesson study meetings and provided invaluable guidance in designing this workshop. Her expertise and thoughtful insights into supporting preservice teachers significantly enhanced our approach to integrating AI tools for task modification. We also acknowledge the 23 elementary preservice teachers who participated with enthusiasm and critical engagement. Their reflections provided the foundation for the insights presented in this manuscript. Finally, we thank our lesson study team colleagues whose interdisciplinary expertise in mathematics education and ESOL instruction enriched this collaborative effort.

AI Transparency and Disclosure

The authors used ChatGPT (GPT-4o free version) as a research tool and object of study in this manuscript. ChatGPT was employed as an instructional tool within the workshop described in this study, where preservice teachers used it to modify mathematical story problems.