Journal of Urban Mathematics Education
March 2025, Vol. 17, No. 2, pp. 11–42

Influence of Language on Multilingual Middle Grades Learners' Mathematical Problem-solving Outcomes

Jodie Cahill
Bowling Green State University

Jonathan D. Bostic
Bowling Green State University

Multilingual learners (MLs) encounter key language features that influence their mathematical word problem-solving processes. Much of the readily available literature surrounding MLs' experiences in the mathematics classroom explores either their content knowledge or how language affects their mathematics achievement. Far less literature has examined problem-solving outcomes among MLs. The present study attempts to fill a gap in the literature by examining seventh-grade MLs' engagement with grade-level word problems that align with mathematics standards and are realistic, complex, and open (Verschaffel et al., 1999). An explanatory mixed-methods approach (Creswell, 2012) was utilized to explain the features underlying ML's problem-solving outcomes. MLs were interviewed via a Retrospective Think Aloud protocol following their engagement with each word problem. Findings indicate that MLs are extremely successful at engaging in sensemaking, a vital aspect of problem-solving, based on their abilities to describe the problem context presented in their own words. Further, words having both mathematical and nonmathematical connotations were utilized by MLs in terms of their nonmathematical connotations, which influenced their mathematical word problem-solving processes. These two findings add to past studies (Gándara & Hopkins, 2010) and highlight a need to integrate English Language Proficiency skills throughout mathematics instruction.

KEYWORDS: Equity, middle grades, mixed method, multilingual learners, problem solving

Words used in a mathematics setting often take on different meanings from their meanings in everyday life, posing a particular challenge for multilinguals (MLs; Bay-Williams & Livers, 2009). Among other things, academic mathematics

Jodie Cahill, cahillj@guilderlandschools.net , completed this research while a graduate student at Bowling Green State University. She is currently a seventh-grade mathematics teacher in Guilderland Central School District. She is interested in connecting DEI initiatives to the content she teaches classroom and promoting such initiatives throughout her district.

Jonathan D. Bostic, bosticj@bgsu.edu, is a professor of mathematics education at Bowling Green State University. He studies mathematical problem solving in K-20 settings as well as assessment across STEM Education contexts. Queries for both Jodie and Jonathan may be sent to the following address: 529 Education Building, Bowling Green, OH 43403.

vocabulary may have a different use in everyday language (i.e., right, table, row) or may be related to but more specific than use in everyday language (i.e., difference, average, reflect; Bay-Williams & Livers, 2009; Bush et. al., 2020). Thompson and Rubenstein (2000) highlight several vocabulary issues to consider when comparing the languages of mathematics and everyday English. One such issue is that "some mathematics words are shared with English and have comparable meanings, but the mathematical meaning is more precise" (Thompson & Rubenstein, 2000, p. 569). Examples of such vocabulary terms include, but are not limited to: divide, difference, limit, similar, average, and edge (Thompson & Rubenstein, 2000). Bay-Williams and Livers note a similar categorization of mathematics vocabulary terms: words that may be related but more specific than used in everyday language (2009). Resources are available to help teachers design instruction that implements clear and concise mathematical language (see Bush et al., 2022; Fuchs et al., 2021, for examples) as a potential intervention in the mathematics classroom. Much of the research outlined above, while telling, is not explicit to MLs' needs. There are language features that affect all students' engagement with mathematical word problems (Bay-Williams & Livers, 2009; Thompson & Rubenstein, 2000), and the present study attempts to add to a limited body of research that focuses specifically on MLs' problem-solving. The purpose of this study is to explore ML's word problem-solving outcomes.

Literature Review

Definition of Terms

The term native-English speaker (NES) is used to characterize those students who speak English as their first language. The World-Class Instructional Design and Assessment (WIDA) uses the term multilingual learners to describe all students who interact in language in addition to English on a regular basis (WIDA, 2020). Thus, the term multilingual learner (ML) has been carefully chosen for this study to promote multilingualism as an asset and to avoid deemphasizing students' other languages. This choice of terminology is intended to be both asset- or strengths-based and inclusive (WIDA, 2020). Other common labels, such as English for Speakers of Other Language (ESOL) and English Language learner (ELL) used for MLs may implicitly be deficit-oriented because they emphasize the language of instruction (English) and deemphasizes students' native languages (Barwell et al., 2017).

Framing Urban Education and its Scholarship

With the intent to inform teaching, learning, and policy in urban mathematics classrooms, the researchers identify themselves as scholars of urban mathematics education (Young et. al., 2022). In identifying as such, the researchers hope to bring attention to issues of power, race, and identity while addressing geospatial concerns through theoretically and empirically sound research (Martin & Larnell, 2013; Welsh & Swain, 2020; Young et. al., 2022). Welsh and Swain (2020) articulate four tenets that help to conceptualize and operationalize urban education, which we draw upon for this study. First, urban education is "dynamic and complex rather than static and monolithic settings, with communities that continue to be shaped by the vestiges of a discriminatory and oppressive past" (p. 97). That is, urban education is something that changes (Wells, 2014). Second, urban education is framed by a range of characteristics, challenges, and context and not necessarily as a dichotomous view of meets or not meets. Historically, urban districts "enroll more low-income, minority, and ELL students" (p. 97) than suburban and rural districts, but those features are just one part of the rich context of urban education. Third, urban education has been and continues to be defined by historical, present, and potential educational inequalities. Fourth, urban education scholarship rejects the narrative of deficit perspectives and highlights the assets of urban communities. This final tenet argues against gentrification interventions and deficit perspectives and instead, recognizes and raises up the assets within urban communities. These four tenets connect to the mission of this journal, recent urban mathematics education research that grounds the present study, and serve as a broad framework for this study.

Martin & Larnell (2013) note that various historical moments have served to create conditions, both inside and outside of mathematics education, that require an *urban* mathematics education. These conditions are not static (Martin & Larnell, 2013), and one such condition involves many urban classrooms facing an influx of transient students (Nevárez-La Torre, 2012). Consistent, accurate information about the demographics of this population is difficult to obtain, due in part as a result of migrant students' high mobility (Nevárez-La Torre, 2012). Since 2016, nearly one million immigrant students (i.e., students living in the USA for less than three years) have been enrolled in K-12 schools (Sugarman, 2023). During that same time, there are approximately five million MLs in K-12 schools across the USA (Sugarman, 2023). Further, urban areas are more likely than rural areas to have students who identify as MLs (Bialik et al., 2018). MLs made up 14% of total public school enrollees in urban school districts in 2015 compared to just 4% in rural areas (Bialik et al., 2018). More specifically, MLs' enrollment in urban school districts ranges from 9.6% in small cities to 16.6% in large cities (Kena et al., 2016). These students are part of migrant families who decide to reside in urban areas with the hopes of both achieving financial stability and accessing better educational opportunities for their children. These students are often identified as MLs, and a particular challenge that teachers in urban settings face is being unequipped to work with MLs who have a transient background and who speak a language for which few instructional resources exist (Nevárez-La Torre, 2012).

The researchers for the present study intend to maintain a strengths-based view of MLs throughout the following research. However, in reviewing the relevant literature, MLs and their achievement were often framed in a deficit light. While contexts differ, Stinson's (2006) review provided some background for this portrayal. Stinson documented the discourse of deficiency, rejection, and achievement surrounding Black students when compared to their White counterparts in order to successfully prove that academia has historically categorized marginalized groups in terms of what they cannot do rather than what they are capable of doing (Stinson, 2006). In discussing the discourse of deficiency, Stinson described how Black students were often characterized broadly in literature as "incapable of measuring up to schools' predetermined goals and objectives and lacking the behavioral and social skills and life experiences to be academically successful" (2006, p. 485). The discourse of rejection routinely focused on the systematic rejection of school and academics by Black students. That is, scholarly works oftentimes framed Black students as rejecting school or academics (Stinson, 2006). Finally, the discourse of achievement framed the achievement gap between Black and White students for reasons such as deficiency or rejection (Stinson, 2006). The researchers for the present study found similar alarming findings when compiling sources for the literature review to support current research: much literature surrounding MLs' mathematics experiences in K-12 education has utilized a deficit-based lens. MLs were often discussed in scholarly works in terms of their deficiencies, their rejection of schooling, and their lack of achievement as compared to their NES counterparts. It is important to make note of such a trend so that the reader understands that the authors do not intend to portray MLs in this light but rather communicate what has been published.

Multilingual Learners: General Context

Historically, MLs have not been provided with the same opportunities to engage fully in the United States of America's (USA) public education system as their NES counterparts (Planas & Civil, 2013). For example, in Arizona, Proposition 203 was passed in the year 2000, which severely restricted ML education (Planas & Civil, 2013). The underlying assumptions of the proposition positioned MLs' native languages and cultures as subordinate and inferior to English (Planas & Civil, 2013). A few years later, the Arizona legislature passed a law that essentially segregated MLs from their NES peers for four hours every day to focus on learning English (Planas & Civil, 2013). This segregation denied MLs access to the core curriculum and the stakes for these issues continuously increase as the population of the USA becomes more diverse over time. The National Education Association (NEA, 2020) projects that one in four USA students will be an ML by the year 2025. One challenge many MLs face with U.S. schools is tests administered in one language (Gándara & Hopkins, 2010; Rodriguez et al., 2020). A second challenge for MLs in USA schools is that they are learning academic English as well as content in all subject areas (Gándara & Hopkins, 2010). At the same time, many are enrolled in English Language Proficiency (ELP) classes instead of subject-area classes taught by a content-focused teacher (Gándara & Hopkins, 2010; Planas & Civil, 2013). Thus, MLs have fewer contact hours with content-area teachers. Although MLs are working to master ELP skills, they do not have the same opportunities to master content from core subjects as their NES peers (Gándara & Hopkins, 2010).

Multilingual Learners: Mathematics Context

MLs' Mathematical Content Knowledge

MLs' language proficiency in both their native and additional languages has been linked to their mathematical performance (de Araujo et. al., 2018). MLs with high language proficiency in both their native and English language categories perform better on mathematical assessments than students with low proficiency in both languages (de Araujo et. al., 2018). This supports the claim that MLs' ELP may correlate with mathematical proficiency and content understanding. An argument can be made that supporting MLs' ELP within mathematics content courses has the potential to positively influence ELP and, in turn, boost MLs' mathematics achievement. However, it is important to note that content knowledge should not be viewed as fully dependent on ELP (McGraw & Rubinstein-Avila, 2009). Effective mathematics teachers of MLs must provide quality mathematics instruction and may need to take actions such as reinforcing the connections between mathematics representations and their meaning, exploring the meanings of mathematical words and objects through speech and other forms of expression, and making conscientious use of visual displays so that MLs have access to pertinent information throughout instruction (Sorto, 2019). Further, Fuchs et. al. (2021) suggests routinely teaching mathematics vocabulary to build students' understanding of the mathematics they are learning, using clear, concise, and correct mathematical language throughout lessons to reinforce students' understanding of important mathematical vocabulary words and supporting students in using mathematically precise language during their verbal and written explanations of their problem-solving. Such instructional strategies help to promote a more equitable learning environment in which MLs have ample access to learning opportunities (Fuchs et. al., 2021; Sorto, 2019).

Language and Mathematics

Promoting students' mathematics growth through instruction is tied to language (NCTM, 2000; 2014). Mathematics instruction that promotes mathematics language development is central to students' learning (NCTM, 2000). Much literature discusses the importance of teaching clear and concise mathematical language to help students effectively communicate their understanding of mathematical concepts (see Bush et al., 2020; Fuchs et al., 2021, for examples). While these sources are sufficient in assisting a general population of students, integrating language into research on mathematics learning is crucial for improving MLs' engagement in mathematics education (Moschkovich, 2010). There have been some studies conducted surrounding MLs' language use in mathematics settings. One such study, which took place in Papua New Guinea, focused specifically on the terms more and less as they were used by elementary-aged MLs (Jones, 1982). Jones described that the word more was consistently mastered by NES students before the word less when used in the same context (1982). However, studies discussed that MLs showed greater strength with the word more in early elementary school and, in a sudden reversal, showed greater strength with the word less in late elementary school (Jones, 1982). MLs were found to be two to four years behind their NES counterparts in mastering contexts for both the words more and less (Jones, 1982), which is consistent with Barrow's (2014) claims that social language is acquired by MLs before academic language. The consequence of such a delay in the setting of the Jones study was that MLs were still struggling to understand the relationship between more and less long after they had been taught the mechanics of addition and subtraction (1982). In short, MLs were performing procedures for which they did not have a robust understanding due to a language issue. Jones concluded that in order to improve the quality of mathematics instruction for those learning mathematics in an additional language, "more needs to be known about those aspects of language which impinge directly on the learning of mathematics, and which present particular difficulties for those learning in their [additional] language" (1982, p. 286).

While there are language factors that affect MLs' engagement in mathematics at the classroom level (Jones, 1982), these factors also impact MLs' performance on high-stakes tests. Martinello (2008) analyzed the linguistic complexity of the Massachusetts Comprehensive Assessment System (MCAS) fourth-grade mathematics items and used think-aloud protocols to gather evidence of comprehension difficulties for Spanish-speaking MLs. MCAS was the high-stakes summative test used by the state, and Martinello focused on the English version of the exam (Martinello, 2008). It was found that items with both complex vocabulary (e.g., low-frequency words, culture-specific nonmathematical vocabulary terms) and sentence structures (e.g., mean sentence length in words, number of prepositional phrases) disfavored MLs compared to NES (Martinello, 2008). Words with multiple meanings posed additional challenges to MLs' reading comprehension (Martinello, 2008). As an example, some students confused *one* as a pronoun with *one* as a numerical item. It was observed that MLs may only know the most familiar connotation of a word with more than one meaning. In these cases, MLs tended to assign the most familiar meaning to such words regardless of context (Martinello, 2008). For example, one item appearing on the MCAS contained the phrase, "The pictograph below shows the amount of money each fourth-grade class raised for an animal shelter" (Martinello, 2008, p. 358). Some MLs who engaged with this item interpreted the verb *to raise* to mean *to increase* or *to rear* animals rather than *to raise funds* (Martinello, 2008). Martinello argued for more studies using think-aloud protocols to investigate how MLs interpret mathematics word problems of varying linguistic complexity (2008).

Another study that focused on MLs' language in a mathematics setting took place in Catalonia, where the immigrant population, thus, the ML population, had increased significantly (Gorgorió & Planas, 2001). In Catalonia, educational administrators believed the only problem MLs faced involving their schooling was not knowing Catalan, the native language (Gorgorió & Planas, 2001). However, Gorgorió and Planas believed that considering only language acquisition when it came to ML achievement was insufficient (2001). They claimed that educators must also consider that the basic use of an additional language did not imply that MLs had linked word meanings to the culture of that additional language (Gorgorió & Planas, 2001). A goal of the Gorgorió & Planas (2001) study was to observe students during mathematics lessons to identify problems regarding language. The researchers found that mathematical difficulties experienced by MLs were not always due strictly to a language obstacle, but to a wider communication obstacle (Gorgorió & Planas, 2001). Behind most social and cultural issues in a mathematics classroom were the meanings attached to situations within the context of its social dynamics (Gorgorió & Planas, 2001). Because words are connected with different meanings for different individuals, being aware of the misunderstandings that could appear in a multilingual situation could help teachers and researchers understand both social and cultural conflicts. For example, a Bangladeshi boy who took part in the study expressed that the symbol for the digit 6 in his native language was very similar to the way 7 was written in Catalan (2001). Therefore, this student's prior experiences with mathematical words and symbols affected his ability to add and subtract (Gorgorió & Planas, 2001).

Elementary-aged students have been the primary focus of studies of how language may affect MLs in a mathematics setting (Gorgorió & Planas, 2001; Jones, 1982; Martinello, 2008). However, Ní Ríordáin and O'Donoghue (2009) conducted research in Ireland with secondary students and the country's two official languages. The study focused on students transitioning from Gaeilge-medium mathematics education to English-medium mathematics education (Ní Ríordáin & O'Donoghue, 2009). All participants took both a mathematics word problem test and a language proficiency test in English. MLs also completed the mathematics word problem test and a language proficiency test in Gaelige in addition to the English tests. Results indicated that performance on mathematical word problems and language proficiency in English were moderately correlated (r=0.48). Further, the relationship between performance on mathematical word problems in English and language proficiency in Gaelige suggested a strong relationship was evident (r=0.65; Ní Ríordáin

& O'Donoghue, 2009). Those with high proficiency in both languages outperformed their monolingual peers on the mathematics test. It is quantitatively clear that language proficiency and mathematics education are related, and multilingualism may enhance mathematics performance on word problems (Ní Ríordáin & O'Donoghue, 2009). Taken collectively, MLs' mathematics achievement has less to do with their age or grade level and more to do with their language proficiency.

MLs and Word Problem Solving

For the present study, we selected Lesh & Zawojewski's (2007) framing of problem-solving; it is a process that requires "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). Given the intent to focus on word problems, the researchers used a framing of mathematical word problems such that they are tasks that promote mathematical problem-solving and are realistic, complex, and open (Verschaffel et al., 1999). Realistic word problems draw upon experiences from a local, regional, or national context. Complex word problems are meant for students to engage in creative and critical thinking while drawing upon multiple aspects of mathematics knowledge. Open word problems can be solved with multiple developmentally appropriate strategies and within instruction (Verschaffel et. al., 1999). Drawing from Verschaffel et al.'s (2000) related research, the researchers leveraged a six-stage problem-solving framework when considering how students engage in mathematical word problem-solving. This framework includes: (1) reading the problem, (2) creating a representation of the situation, (3) constructing a mathematical representation of the situation, (4) arriving at a result from employing a procedure on the representation, (5) interpreting the result in light of the situational representation and (6) reporting the solution within the problem's context (Verschaffel et. al., 2000). One key factor of mathematical word problem-solving is sensemaking, which is when students develop an understanding of a situation or context by connecting it with existing knowledge (Matney et al., 2022; Verschaffel et al., 2000). Students who struggle to comprehend the relationship between word problem text and a representation of the situation have been shown to then struggle to create a mathematical representation of the situation (Matney et al., 2022; Pape, 2004; Verschaffel et. al., 2000). That is, in order to be an effective mathematics problem solver, students must demonstrate sensemaking (Matney et al., 2022).

The bulk of literature on MLs experiences in the mathematics classroom explores either their content knowledge (de Araujo, 2018; McGraw & Rubinstein-Avila, 2009; Sorto, 2019) or the role language plays in their mathematics achievement (Gorgorió & Planas, 2001; Jones, 1982; Martinello, 2008; Moschkovich, 2010; Ní Ríordáin & O'Donoghue, 2009). While issues of language arise at each stage of

Verschaffel et. al.'s (2000) problem-solving framework, these are not explicitly linked to MLs' achievement. In short, little data are available regarding how MLs perform during word problem-solving. Some have recommended that more scholarship should carefully explore MLs' problem-solving abilities (e.g., Barwell et. al., 2017). Although research shows that MLs may develop procedural fluency equivalent to their NES peers, they may not have developed the oral language to accurately and consistently solve word problems (Sanford et. al., 2020). Word problems are widespread in mathematics classrooms and have particular language features that may be challenging for MLs (Barwell et. al., 2017). Just as there is a link between mathematics and language proficiency, there is a relationship between language proficiency and performance on word problems (Barwell et. al., 2017; Ní Ríordáin & O'Donoghue, 2009). The relationship between language and word problem-solving can be influenced by aspects including but not limited to: (a) language in which the word problems are presented, (b) word problem syntax (Martinello, 2008), (c) words that have both mathematical and nonmathematical connotations (i.e., table, mean, and plane; Bay-Williams & Livers, 2009; Martinello, 2008; Thompson & Rubenstein, 2000), and (d) students' level of reading comprehension (Barwell et. al., 2017). Thus, these factors indicate that word problem-solving is impacted by numerous aspects related to language, which is a heightened issue for MLs.

Synthesis

The present study attempts to examine how MLs engage in mathematical problem-solving within the context of mathematics found in their instructional standards. Additionally, prior research surrounding the effects of language on MLs' engagement in mathematics content beyond elementary contexts is being extended. Elementary-aged MLs (Gorgorió & Planas, 2001; Jones, 1982; Martinello, 2008) have been a greater focus when it comes to MLs' mathematical word problem solving, and there are fewer studies at the middle (McGraw & Rubinstein-Avila, 2009) and secondary levels (Ní Ríordáin & O'Donoghue, 2009). Furthermore, these studies of the relationship between language and mathematics achievement for MLs have more to do with arithmetic word problem solving as opposed to word problems that are realistic, complex, and open (Verschaffel et. al., 1999). Thus, this study aims to fill a gap in the literature by exploring key language features that influence MLs word problem-solving through the lens of Verschaffel et al.'s (1999) characterization of realistic, complex, and open word problems and the related problem-solving framework (Verschaffel et al., 2000). The framework of realistic, complex, and open-word problems has not necessarily been a focus within past studies of ML's problem-solving. The research question for the current study is: What are key feature(s) influencing seventh-grade MLs' mathematical word problem-solving?

Method

Research Design

The present study utilizes an explanatory mixed-methods approach (Creswell, 2012). The aim of this study is to examine features underlying ML's problem-solving outcomes. MLs were cognitively interviewed by utilizing a Retrospective Think Aloud (RTA) procedure (Guan et al., 2006), which requires that students solve realistic, complex, and open tasks (Verschaffel et al., 1999). Following the completion of the task, MLs were asked to explain to the researchers the step(s) they took in achieving their answer, among other things. Purposeful homogenous sampling (Creswell, 2012) was utilized to select participants for this study so that all participants shared the characteristic of being or previously being a ML.

Setting

Southwest School was located in a large urban district of the Southwestern USA. One teacher from Southwest School referred to as Ms. Smith was an NES mathematics teacher. Ms. Smith had 133 students across five seventh-grade mathematics classes. Mathematics instruction was exclusively delivered in the English language per district guidelines. However, Ms. Smith allowed her students a choice in what language they used when talking with one another about mathematics during classroom instruction. She expressed her primary goal was for students to develop grade-level mathematics knowledge and skills. As such, she allowed students to use English or another language when they communicate with peers who speak the same language. Ms. Smith communicated that most students spoke in English when working in groups on a typical day. However, some MLs used Spanish, the dominant language other than English in this setting, if they were working with a close friend who also spoke Spanish or if they were processing their thinking aloud (i.e., talking to themselves).

Participants

Students from Southwest School in grade seven participated in this study. All students who were identified as MLs from this setting spoke Spanish as their native language. Most students who participated in the present study were enrolled in a formal ELP program offered at Southwest school. Some identified MLs had graduated from a formal ELP program, and these students also participated in the study as English was not their first language. Twenty-eight of Ms. Smith's students were interviewed via the RTA protocols.

Table 1

Demographic information of participants

Gender Identity		
Female	Male	Prefer not to Say
14 (50%)	13 (46%)	1 (4%)
Mathematics Ability Level		
Above Average	Average	Below Average
10 (36%)	8 (28%)	10 (36%)
Whether Student is on IEP or 504		
Yes*		No
3 (11%)		25 (89%)
Race/Ethnicity		
Hispanic	Native American	Mixed Race
24 (86%)	2 (7%)	2 (7%)

*Of those with an IEP or 504, two students have identified reading disabilities and one student has an identified math disability.

Data collection

Instruments

Three tasks aligned with the open, complex, and realistic framework were selected. They were adapted from released items taken from the grade seven Problem Solving Measure (PSM; Bostic et al., 2024; Bostic et al., 2017). The PSM is co-owned by two universities and available to researchers, evaluators, and school personnel. Published literature indicates that all items were reviewed by a panel of grade-level mathematics teachers, mathematics educators, mathematicians, and a diverse bias panel during the validation process (Bostic et al., 2024; Bostic et al., 2017). Table 2 details both the word problem as well as the Common Core State Standard for Mathematics (CCSSM) content standard aligned to each task. Each item has a Flesch-Kincaid score below seven, indicating that each item's readability score is at or below a typical seventh-grade student's reading level grade level. *Nature Park Task,* a grade six-item from the PSM (Bostic & Sondergeld, 2015), was utilized to help students acclimate to the RTA.

Figure 1

PSM Items and Alignment with Standards

Item 1	
Nature Park Task	A group of 120 people were waiting for a boat to take them on a trip through a nature park. The boat can carry 14 people on each trip. After several hours, everyone in the group of 120 people had gone through the nature park. What is the fewest number of trips made by the boat to carry this group of people?
Content Standard	**6.RP.3.4** Use ratio and rate reasoning to solve real-world and mathematical problems, e.g., by reasoning about tables of equivalent ratios, tape diagrams, double number line diagrams, or equations.
Item 2	
Women's Height Task	A women's high school volleyball team and a women's high school basketball team each have eight players. Players' heights are shown in the table. What is the difference between the average heights of the teams? In this problem, average is the mean.
Content Standard	**7.SP3.1** Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability. *For example, the mean height of players on the basketball team is 10cm greater than the mean height of players on the soccer team, about twice the variability (mean absolute deviation) on either team; on a dot plot, the separation between the two distributions of height is noticeable.*
Item 3	
Water Tower Task	A water tower contains 16,880 gallons of water. Each day half of the water in the tank is used and not replaced. This process continues for multiple days. How many gallons of water are in the tower at the end of the fourth day?
Content Standard	**7.EE.1** Apply properties of operations as strategies to add, subtract, factor, and expand linear expressions with rational coefficients

Students selected a pseudonym and demographic information prior to engaging in the RTA. Ms. Smith provided information about participants' mathematics ability level based upon past classroom and assessment performance, as well as whether a student received services for a disability.

Retrospective Think Aloud Protocol: Context

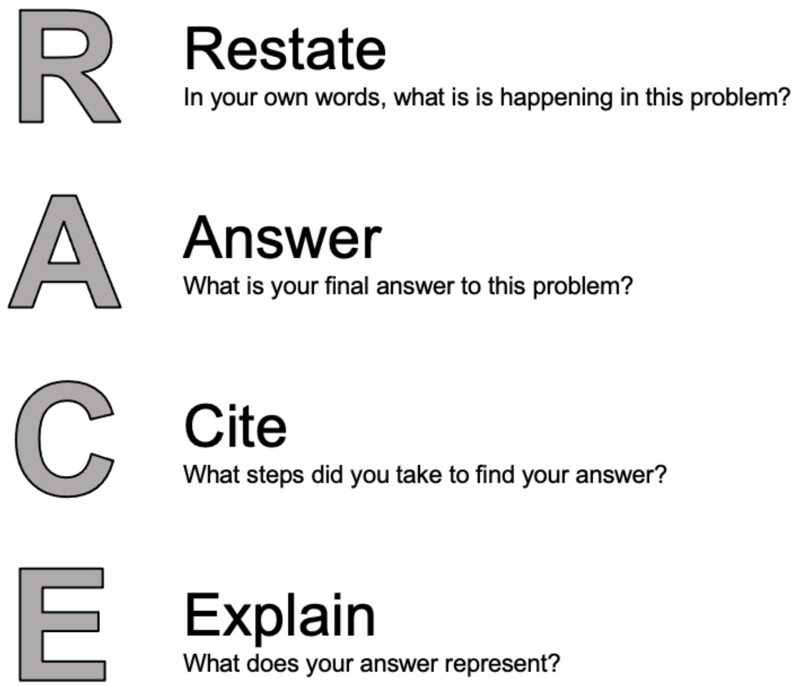
Think-aloud interviews are used to observe, define, and measure the contents of students' minds as they work on educational assessments, especially those focused on problem-solving (Bostic, 2021; Bostic et al., 2021; Leighton, 2017). The procedure used to collect data regarding students' problem-solving processes was a RTA protocol (Guan et al., 2006). In brief, students were asked to complete the tasks. After they indicated that they had completed the task, then they were asked to verbalize their processes (Guan et al., 2006). The most fundamental aspect of RTAs is that students feel encouraged to talk about what they did, with specific attention to their mental processes or performance during problem-solving (Guan et al., 2006). RTAs were chosen as opposed to typical think alouds for two reasons: (1) MLs' teachers indicated that asking students to think aloud and do mathematics work simultaneously is more cognitively taxing than they typically experienced in class, and (2) RTAs give students an opportunity to focus on doing mathematics rather than accessing necessary academic English language to verbally communicate to the research team about their approach. All data collection was videotaped.

Retrospective Think-Aloud Protocol: Procedure

Prior to conducting the study, guidance was sought on ways to structure the RTAs for MLs from three terminally-degreed scholars at universities who served as an expert panel. Their expertise spanned three related areas: Reading, Multilingual Learners, and Special Education. The experts communicated agreement that a discussion protocol would support students' abilities to navigate multiple languages as well as mathematics content and its academic language. This decision is further backed by research: Moschkovich (2010) describes that, historically, mathematics education researchers who address language issues have used work from fields outside of mathematics education to inform research on the relationship between language and mathematics. Discussion protocols support students' access to rich mathematics communication and can be used to support student explanation and engagement with mathematics vocabulary and math-specific word meanings in context (Buffington et al., 2017). It was ideal that Southwest School students used the RACE (see Figure 2) protocol during their English Language Arts class. At the beginning of each RTA, participants noted that they were familiar with the RACE protocol, which is discussed later in this section. A card with the RACE protocol was provided to students during the RTA.

Figure 2

RACE Protocol Used During Data Collection.



The RTA process included three steps: (a) Read the problem, (b) Solve the problem, and finally (c) discuss the problem using the RACE protocol. Students were encouraged to utilize a calculator and write and/or draw on the page if they chose to do so. After students completed each problem, the researchers directed students to the RACE protocol. They took notes as students talked. If students did not clearly address a component of RACE, then the researcher asked them to do so before moving on. This process repeated for all four tasks.

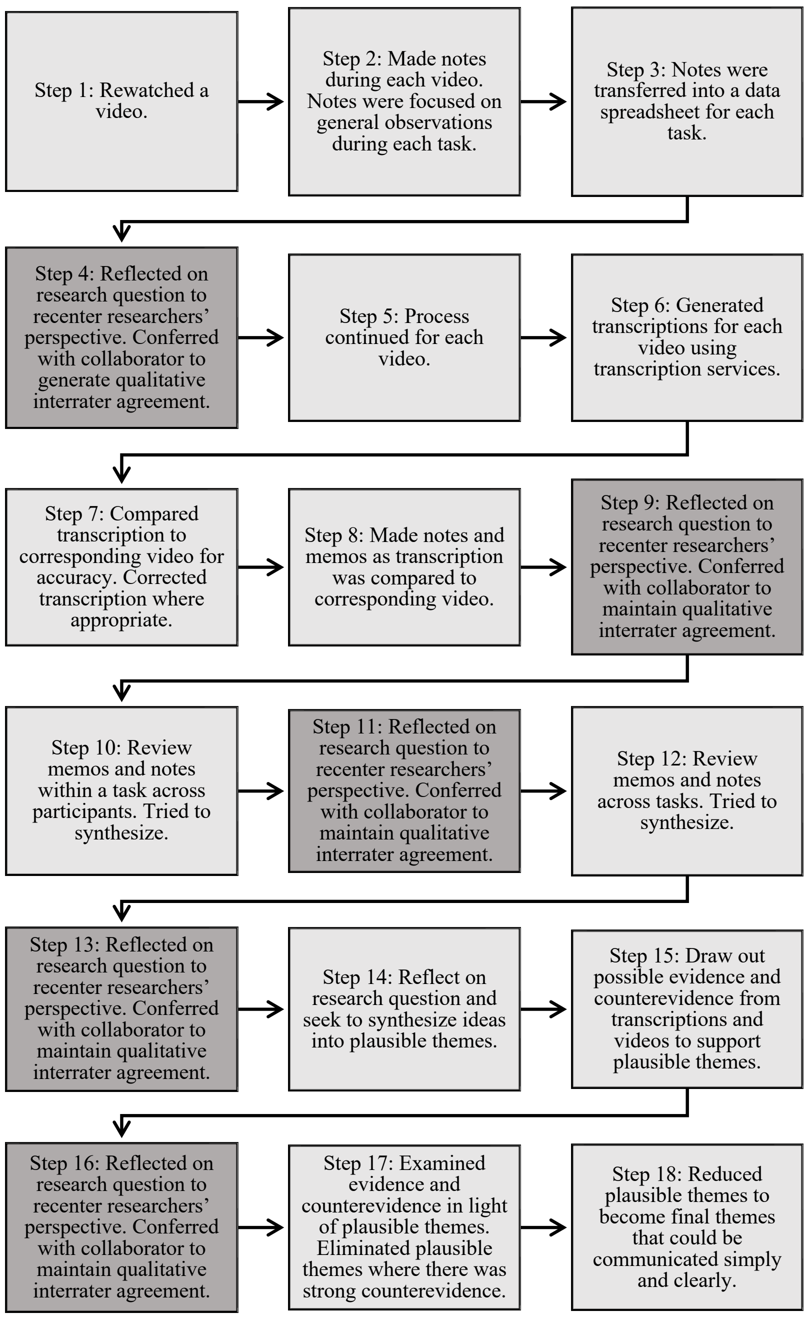
Data Analysis

Qualitative inductive analysis was used to draw themes from the data (Creswell, 2012; Strauss & Corbin, 1998). A multistage process was developed to maintain fidelity within each researcher's analysis as well as across researchers' analysis, which is shown in figure 3. Checks and balances were performed at multiple steps (i.e., steps four, nine, eleven, thirteen, and sixteen) across the research team to ensure there was interrater reliability during data analysis. Checks and balances constituted opportunities for the two researchers to evaluate their process and outcomes. An example is shared from step nine. Prior to this step, the two researchers had worked independently and at this step, came together to share their work products. Disagreements were discussed until a consensus was reached. Triangulation (Miles et al., 2014) tended to leverage different data sources (i.e., different participants) to help the researchers reach consensus through corroboration. Ultimately, triangulation became "a way of life" (Miles et al., 2014, p. 300) as a means to foster strength of claims. Qualitative interrater reliability was built to consider "whether the process of the study is consistent, reasonably stable over time and across researchers" (Miles et al., 2014, p. 312). The checks and balances during the process helped to foster and retain robust qualitative interrater reliability (Miles et al., 2014). Some of Miles et al. (2014) suggestions that the researchers followed to promote qualitative interrater reliability included: Using comparable data collection protocols, having multiple interrater checks during data analysis, and reviewing data for potential bias. Taken collectively, our multistage qualitative data analysis process was intended to build trustworthiness, maintain strong coherence, and to limit bias.

The problem-solving framework (Verschaffel et al., 2000) overlaid the analysis and became more prominently considered during the checks and balances. A goal during the checks and balances stages was to reflect using the problem-solving framework. Discussions about the memos, noticings, and/or potential themes during steps four, nine, eleven, thirteen, and sixteen included reflections on how the framework the data might reify different facets of the Verschaffel's six step problem-solving framework. For example, we considered how students' language might demonstrate their actions during situation modelling or it was better characterized as part of the mathematical modeling stage. Again, a goal during these checks and balances was to reflect back to the problem-solving framework.

Figure 3

Multistage Qualitative Inductive Analysis Process



Findings

Two themes emerged from the Qualitative Inductive Analysis process: (1) MLs were successful in sensemaking by creating accurate situation models from the contexts of the word problems, and (2) MLs tended to utilize mathematical words with multiple meanings in terms of their nonmathematical connotations. The authors have chosen to first situate these qualitative findings quantitatively in terms of MLs' performance to better understand these themes. MLs performed similarly to the general population of students who took the PSM assessment.

ML Performance

The task sets given to MLs were scored dichotomously, which allowed problem-solving to be more fully contextualized with the qualitative data. Table 2 summarizes how many participants achieved a solution for each item.

Table 2

MLs' Overall Performance on Task Sets

	Correct	**Incorrect**
Nature Park Task	3	25
Women's Height Task	3	25
Water Tower Task	7	21
Total	13	71

The mean score on the task set was a 0.46 out of 3 possible points. On average, MLs earned a 15.3% on the task set. This score is consistent with the normed performance for the 15-item PSM, which also tends to be low (Bostic et al., 2017). MLs' quantitative outcomes for the present study are consistent with generalized findings across NES participants on similar items.

Sensemaking and Creating Situation Models

It was evident that MLs were able to engage in the "restate" step of the RACE protocol much of the time. Success during the restate step of the protocol was operationalized as accurately describing the problem situation. Each of the 28 participants engaged with three items, thus, there were a total of 84 ML interactions with items. Of these 84 interactions, a full understanding of the items' contexts was demonstrated 81 times. That is, MLs were successful 96% of the time at engaging in the "restate" step of the RACE protocol. MLs communicated evidence that they created accurate situation models orally throughout the RTA process, which is reified through two examples. One ML, Billie, successfully restated what was happening in *Water Tower Task*. Figure 4 below details the conversation Billie had with one researcher, which exemplifies MLs' sensemaking abilities.

Figure 4

Billie's Transcription

Researcher	Okay, so what do you think is going on in this problem?
Billie	Um, there's a water tower that contains 16,880 gallons of water and each day, um, they use half of the water. For four days. And I'm trying to find how much water they would have at the end of the fourth day.
Researcher	Okay, so what would you do mathematically to solve that?
Billie	Um, to make this number [points to 16,880] smaller.
Researcher	Okay.
Billie	Well, I have to make this number smaller four times since it's the end of the fourth day.
Researcher	Okay, so you have to make 16,880 smaller four times?
Billie	Mhm.
Researcher	So, do you know how to make it smaller?
Billie	No.

Not only does Billie use her own words to describe what is happening in the problem, but she continues to demonstrate her understanding of the problem context when she mentions that she will have to make the number "smaller four times" because the water levels are decreasing for four days.

Adrian was another ML who exemplified this sensemaking theme. When engaging with *Nature Park Task*, Adrian successfully restated what was happening in the problem. Figure 5, below, details his conversation with one researcher.

Figure 5

Adrian's Transcription

Adrian	I'm stuck on this one. I don't know what to do.
Researcher	Okay, so let's go through this [points to the RACE protocol prompts]. Can you tell me in your own words what's happening in this problem?
Adrian	There's 120 people trying to go on a trip.
Researcher	Okay, where are they going?
Adrian	A nature park.
Researcher	Alright, what else is going on?
Adrian	The boat can only carry 14 people on each trip.
Researcher	Okay, so what's the question asking you?
Adrian	It says that after several hours, the whole group had gone through the nature park and to find the fewest number of trips made by the boat.
Researcher	Okay, so do you know what that question is asking you to do math-wise?
Adrian	No.
Researcher	So, you understand what's going on in the problem, you're just not sure what math to do?
Adrian	Yeah.

In this case, the researcher explicitly asked Adrian whether it was the problem context or the mathematics that was inhibiting him from entering into a solution strategy. Adrian indicated that he was not sure what steps to take mathematically even though he was able to create and communicate an accurate situation model of *Nature Park Task*.

In the instances where MLs did not successfully restate what was happening in the problem, then it was common for them to reread parts of or the entire item text word-for-word aloud to the researchers. However, MLs were overwhelmingly successful in creating accurate situation models for the items presented to them. This is indicative of MLs' abilities to engage in sensemaking, a vital facet of problem solving. Billie and Adrian, among most others, were able to use their own words to restate what was happening in each item rather than rereading the item text to the researchers.

Mathematical v. Nonmathematical Connotations

It was also evident that MLs understood language within the items that had both mathematical and nonmathematical connotations in terms of their nonmathematical connotations. Of the 84 interactions participants had with the items, MLs utilized a mathematics vocabulary word in terms of its nonmathematical connotation 26 out of 84 possible times (i.e., 31% of the time). As an example of this theme, the word average in *Women's Height Task* was frequently utilized as meaning typical or common by MLs, even though the problem states that participants should use the mathematical definition of arithmetic mean. Figure 6 details a conversation between Delilah and the researcher exemplifying this theme.

Figure 6

Delilah's Transcription

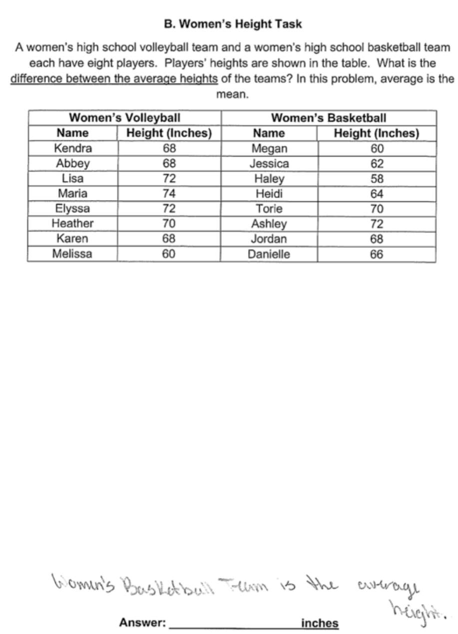
Delilah	So am I supposed to, like, choose the one who has the average height? And then write it down?
Researcher	Whatever you think to do. So, tell me, what do you think the difference is? [pause] Have you noticed any differences?
Delilah	Yeah.
Researcher	Okay, so what differences have you noticed?
Delilah	This one has more 70s and more of the higher. [points to the volleyball team]
Researcher	Okay, so that one has more 70s?
Delilah	Yeah.
Researcher	So what does that mean to you?
Delilah	These ones are taller. [points to the volleyball team]
Researcher	So you think the volleyball team is taller than the basketball team?
Delilah	Yeah. So I think the basketball team is the average height.
Researcher	And why do you think that?
Delilah	Because they're more like, smaller.

Researcher	So the basketball is average? Because they're not as tall? Is that what you're saying?
Delilah	Because the inches aren't as high.

Delilah's final answer to the Women's Height Task reads, "Women's Basketball Team is the average height," shown below. Through this conversation, Delilah communicates that individuals who are tall do not fit into what is commonly considered to be an average height. She concluded that the basketball team was average, utilizing the word in terms of its nonmathematical connotation.

Figure 7

Delilah's Work



B. Women's Height Task

A women's high school volleyball team and a women's high school basketball team each have eight players. Players' heights are shown in the table. What is the difference between the average heights of the teams? In this problem, average is the mean.

Women's Volleyball		Women's Basketball	
Name	**Height (Inches)**	**Name**	**Height (Inches)**
Kendra	68	Megan	60
Abbey	68	Jessica	62
Lisa	72	Haley	58
Maria	74	Heidi	64
Elyssa	72	Torie	70
Heather	70	Ashley	72
Karen	68	Jordan	68
Melissa	60	Danielle	66

Women's Basketball Team is the average height.

Answer: ____________ **inches**

The term average was not the only word that appeared in *Women's Height Task* with dual meanings. Another participant, Angel, utilized the word difference in a nonmathematical way throughout his RTA interview. This term did not trigger him to subtract, rather, note observable differences between the two teams. His conversation with the researcher, which further exemplifies the first theme, is detailed in Figure 8, below.

Figure 8

Angel's Transcription

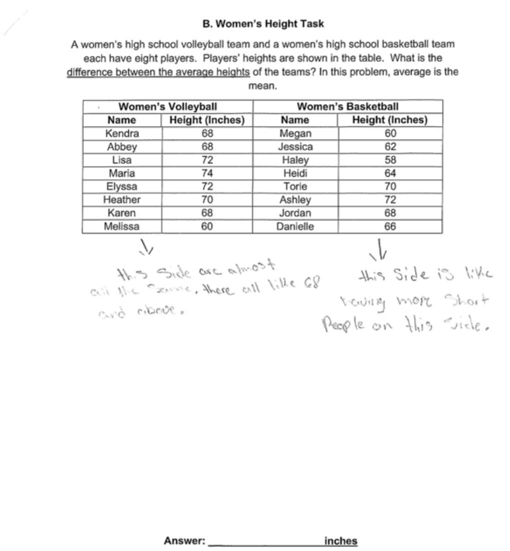
Angel	I kind of didn't really get it. Like, I didn't know if it was supposed to be doing math or it was like you're supposed to write like the differences.
Researcher	Okay, so observe the differences between the teams versus doing some math out?
Angel	Yeah.
Researcher	Okay, so what did you write?
Angel	I just put for the left side [points to volleyball players' height column], this side are almost all the same. They're all like 68 and above.
Researcher	Okay, they're all 68 and above. What do you mean by they're kind of the same?
Angel	Like, because this says average heights, like the heights [on the volleyball team] don't go down. It's just like 68. We have one 60. But this one [basketball team] has 66, 60, 62, 58. Like lower numbers.
Researcher	Okay, so basketball has lower numbers and volleyball has higher numbers. Is that what you're saying?
Angel	Yeah.
Researcher	Okay, have you heard the word average in a math class before?
Angel	Yes.
Researcher	Do you know what it means?
Angel	Sometimes. But sometimes I really don't get it.

Researcher	Okay. Have you heard of the word difference in a math class before?
Angel	Difference is like, isn't that like differences of inches, or, kind of like that?
Researcher	So what does the word difference mean to you?
Angel	Difference means like if someone, like, or I could just use something simple, like a cool kid and a not cool kid. That's what I think difference is.

Angel utilized the nonmathematical connotation for the word difference, which contributes to his wonderings as to whether he was supposed "to be doing math" (Angel). Angel understood the word difference in terms of observable differences between two groups rather than the idea of difference as a result from subtraction. His written work for *Women's Height Task* is shown below. He wrote on the left, "this side are almost all the same. there all like 68 and above." On the right, "this side is like having more short people on this side" is written, further exemplifying that Angel is noting differences in characteristics rather than subtracting.

Figure 9

Angel's Work



B. Women's Height Task

A women's high school volleyball team and a women's high school basketball team each have eight players. Players' heights are shown in the table. What is the difference between the average heights of the teams? In this problem, average is the mean.

Women's Volleyball		Women's Basketball	
Name	Height (Inches)	Name	Height (Inches)
Kendra	68	Megan	60
Abbey	68	Jessica	62
Lisa	72	Haley	58
Maria	74	Heidi	64
Elyssa	72	Torie	70
Heather	70	Ashley	72
Karen	68	Jordan	68
Melissa	60	Danielle	66

this side are almost all the same. there all like 68 and above.

this side is like having more short people on this side.

Answer: ________________ inches

Women's Height Task was not the only item given to MLs where mathematical words were utilized in terms of their nonmathematical connotations. Valerie understood the word half in *Water Tower Task* to mean splitting something up. Similar to Angel not immediately thinking of subtraction when seeing the word difference, half did not suggest division by two to Valerie. At times during a conversation, half was utilized imprecisely, meaning to split something into two parts, but the parts may not be equal. Because this nonmathematical connotation was Valerie's understanding of the word, it was difficult for her to enter into a viable solution strategy. Her conversation with one researcher is detailed in Figure 10.

Figure 10

Valerie's Transcription

Valerie	How much am I supposed to take away?
Researcher	[rereading the item] Each day, *half* of the water in the tank is used and not replaced.
Valerie	So how much is that?
Researcher	What it says in the problem: half of the water.
Valerie	So 16,880?
Researcher	What does half mean?
Valerie	Like, split up, kinda, like, in half. I don't really know.
Researcher	Yeah, let's move to a different one because I think what you're telling me is that with that one, you're not sure what half means.
Valerie	Yeah.

Valerie ultimately provided no answer for *Water Tower Task* because she did not consider half as meaning to divide by two. Therefore, she was unable to enter into any solution strategy. It is evident from the three exemplars provided that words such as average, difference, and half, which all have both mathematical and nonmathematical connotations, are being utilized in terms of their nonmathematical connotations by seventh-grade MLs in this study.

The data captured from this mixed-method study indicate that MLs were successful in creating accurate situation models when presented with realistic, complex, and open word problems. Words having both mathematical and nonmathematical connotations may be challenging students from entering into a viable solution strategy that could potentially solve the word problem. This could be a possible explanation for MLs' problem-solving success rate. Utilizing these words in a nonmathematical fashion did not seem to inhibit MLs' understanding of the problem contexts. Rather than situational contexts presented in the items, language appeared to be the most prominent feature that influenced MLs' problem-solving abilities.

Discussion

The research question for the present study was: What are key feature(s) influencing seventh-grade MLs' mathematical word problem-solving? The findings indicate that (a) the situation contexts presented likely did not inhibit MLs' ability to communicate a representation of the problem's situation, and (b) mathematical words that also have nonmathematical connotations may have influenced MLs' mathematical word problem-solving.

Influence of Situation Contexts

One key piece of Verschaffel et. al.'s (2000) problem-solving framework includes creating a representation of the situation presented. Students engage in sense-making while developing a model of a word problem's situation (Matney et al., 2022; Pape, 2004). Prior literature has shown that creating an accurate situation model for a mathematical word problem is a common struggle among monolingual students (Matney et al., 2022; Pape, 2004). However, the participants in the present study were successful at creating accurate situation models. That is, in 96% of cases, participants were able to mentally create and orally express an accurate representation of the situations through the RTA interview process. However, they were unsure about a mathematical strategy to reach a solution and ultimately showed difficulties with implementing viable solution strategies. This difficulty is consistent with difficulties faced by monolingual students. It has been shown that students who struggle to create a mathematical representation of a given situation then struggle to arrive at a mathematical solution (Matney et al., 2022; Pape, 2004; Verschaffel et. al., 2000; Yee & Bostic, 2014). Many resources are available to help students bridge the gap between stages two (creating a situation model given a word problem) and three (creating a mathematical model given a word problem) of Verschaffel et. al.'s (2000) six-stage problem-solving framework (e.g., Bay-Williams & Livers, 2009; Bush et. al., 2020; Fuchs et. al., 2021). In addressing the research question as to the key features influencing seventh-grade MLs' word problem-solving, the situational contexts provided in the given items appeared not to inhibit MLs' abilities to restate what was happening in the problem.

Because middle school-aged MLs were able to communicate their development of accurate situation models 81 out of 84 times, the researchers do not necessarily consider realistic, complex, and open-word problems, like those in this study, to serve as an obstacle in their problem-solving performance. Specifically, because the word problems presented to MLs were realistic, they were able to engage in sensemaking by developing an understanding of the given situation and connecting it with existing knowledge (Matney et al., 2022). The participants from the present study were MLs, as they indicated that they spoke a language other than English either at school, outside of school, or both. Ms. Smith also indicated that each of the participants were either currently enrolled or had exited a formal ELP program provided by Southwest School. The findings of the present study have the potential to be generalized to other middle school MLs in similar situations: students from the present study were extremely successful at understanding the problem context presented, which is a key feature of mathematical problem solving (Bostic et al., 2016; Matney et al., 2022; Verschaffel et. al., 2000). One important scholarly outcome is that this study provides evidence that MLs engage in rich mathematical problem-solving much like their monolingual peers have shown with open, realistic, and complex word problems (see Bostic et al., 2016; Matney et al., 2022; Yee & Bostic, 2014). Thus, this scholarship starts to build evidence that others might take up when exploring MLs' mathematical problem-solving. With the substantive increase in MLs found in USA schools (see Bialik et al., 2018; Kena et al., 2016), it is critical to retain an asset-based view on MLs successes and to revisit extant scholarship that has historically not included MLs, like problem-solving research. To that end, our burgeoning work is meant to ignite further intersectional scholarship that includes mathematical problem-solving and MLs.

Influence of Language on MLs' Mathematical Problem Solving

The most prominent feature that influenced MLs' mathematical word problem-solving processes was the appearance of words that had both mathematical and non-mathematical connotations in the item text. Mathematical words appeared on the tasks that fell into words shared with English and have comparable meanings, but the mathematical meaning is more precise (Bay-Williams & Livers, 2009; Thompson & Rubenstein, 2000). The current study's findings are consistent with Martinello's (2008) findings: MLs were likely to assign the most familiar meaning to a word regardless of the context. This claim is supported by the fact that MLs utilized a non-mathematical connotation for a mathematics word that appeared in the item text 31% of the time. Further, Delilah, Angel, and Valerie all provided instances of this finding. Social language develops much earlier than academic language (Barrow, 2014;

Jones, 1982), which also could have been a factor in MLs' assignment of meaning to certain words. It is hypothesized that these meanings (e.g., utilizing the word average to mean typical or common rather than to find the arithmetic mean) were most familiar to MLs (Martinello, 2008), which is likely why meanings were assigned in this way.

MLs' language proficiency in both their native and additional language categories has been linked to performance on mathematical assessments (de Araujo et. al., 2018; Martinello, 2008; Ní Ríordáin & O'Donoghue, 2009). It may be concluded that the participants' utilization of mathematical words in terms of their nonmathematical connotations is indicative of their ELP as it relates to academic, mathematical language. Words with more than one meaning may be a feature that influences not only mathematics assessment performance (de Araujo et al., 2018; Martinello, 2008) but MLs' mathematical problem-solving (Barwell et al., 2017; Ní Ríordáin & O'Donoghue, 2009). The present study builds upon past studies of MLs' mathematical problem-solving with further evidence that words with more than one meaning influenced participants' outcomes on open, realistic, and complex grade-level word problems. Such findings build upon past problem-solving research that has focused on issues of language (e.g., Pape, 2004) and adds to current scholarly and practical discussions around MLs.

The appearance of words with both mathematical and nonmathematical connotations has previously been linked to MLs' word problem-solving achievement (Barwell et al., 2017). Because participants performed at a score of 15.3% on the given items, it is likely that their understanding of mathematical words in terms of their nonmathematical connotations may have inhibited them from using a viable solution strategy. Almost one-third of the ML interactions with items (i.e., 31%) were indicative of this behavior. The present study adds to the current body of literature on MLs and word problem-solving by highlighting that word problem-solving is influenced by language, which is a heightened issue for MLs (Barwell et al., 2017; Ní Ríordáin & O'Donoghue, 2009). When students utilize language nonmathematically, they are likely to struggle with constructing a mathematical model of the situation (Verschaffel et. al., 2000) and later employing a mathematical strategy. This was found to be true even though participants were able to construct situation models and thus engage in sensemaking from the items presented. Put simply, if students are unable to interpret the language from the word problem that is presented as mathematical language, then they are unable to solve the word problem, which has been seen in studies with monolingual students (Matney et al., 2022; Pape, 2004). Resources exist to help teachers bridge the gap between mathematical and nonmathematical connotations of mathematical vocabulary (see Bush et al., 2020; Fuchs et al., 2021 as examples).

Prior research (Jones, 1982; Martinello, 2008; Moschkovich, 2010) has called for more studies to be conducted regarding how language affects MLs' abilities to both interpret mathematics and communicate mathematically. Many researchers have heard and sought to answer these authors' calls. Words that have more than one meaning have been shown to affect mathematics students' outcomes (Barwell et. al., 2017; Bay-Williams & Livers, 2009; Bush et. al., 2020; Fuchs et. al., 2021; Martinello, 2008; Thompson & Rubenstein, 2000), and the findings of the current study concur. It can be concluded from the present study's findings that words with both mathematical and nonmathematical connotations are a substantive feature that influences middle school MLs' word problem-solving. In every case where participants regarded one or more mathematical words in terms of their nonmathematical connotation(s), they did not arrive at a correct solution for that item. While many readily available studies have noted that words with both mathematical and nonmathematical connotations may affect students' mathematics performance (Bay-Williams & Livers, 2009; Bush et al., 2020; Fuchs et al., 2021; Thompson & Rubenstein, 2000), they do not specify the impacts such words may have on MLs. Further, they do not specify the impacts such words may have on MLs' problem-solving with word problems that are realistic, complex, and open. The present study extends current literature by highlighting one specific language feature, words that have both mathematical and nonmathematical connotations, as having an influence on MLs' mathematical word problem-solving.

MLs' Engagement in Realistic, Complex, and Open Word Problems

Barwell et. al. (2017) suggests that more research be conducted on MLs' problem-solving outcomes. A significant body of literature is readily available on MLs' content knowledge (de Araujo, 2018; McGraw & Rubinstein-Avila, 2009; Sorto, 2019) and the role language plays in mathematics achievement (Gorgorió & Planas, 2001; Jones, 1982; Martinello, 2008; Moschkovich, 2010; Ní Ríordáin & O'Donoghue, 2009). The present study attempts to fill a gap by observing MLs' work on word problems, specifically those that were aligned to content standards and are characterized as being realistic, complex, and open (Bostic & Sondergeld, 2015; Bostic et al., 2017; Verschaffel et. al., 1999). Similarly, some problem-solving research (Martinello, 2008; Jones, 1982) with MLs has focused more on arithmetic word problems. Previous studies (Bostic et al., 2016; Matney et al., 2022; Verschaffel et. al., 1999; Yee & Bostic, 2014) have explored monolingual students' problem-solving outcomes with open, realistic, and complex tasks, but less research is available as to how MLs, especially middle-school aged MLs, engage in such. The present study's findings show that there is little variation between how MLs perform on realistic, complex, and open word problems compared to how monolingual students performed.

Limitations Of the Present Study

The present study focuses primarily on seventh-grade MLs' problem-solving when engaging with realistic, complex, and open mathematical word problems. The items used in this study stem from Developing and Evaluating Assessments of Problem-Solving (NSF#1720646, #1720661) and Developing and Evaluating Assessments of Problem-Solving – Computer Adaptive Testing, a larger assessment development project (NSF #2100988; #2101026). The items were reviewed by a diverse bias panel of scholars, which included MLs, to ensure the contexts are general enough to be relatable to most if not all, students. In addition, the items were given to a panel of diverse students, separate from the present study, including MLs, for feedback to confirm the conjecture that the item contexts were relatable and limited in bias. Feedback from the bias panel and students confirmed the conjecture that the items showed limited bias and contexts were relatable. With that in mind, it may have been more likely for participants of the present study to be able to create accurate situation models from the given items because they were reviewed by student and scholarly panels. This creates a potential limitation in that word problems that have not gone through such an extensive and thorough process of limiting bias may not yield the same findings. Further research should be conducted focused on MLs' word problem-solving outcomes like the word problems seen in textbooks and/or common curricula. Another area for future research is to take a more fine-grain approach to examine whether there are different levels of situation modeling, which might be an indication of varying degrees of sensemaking. A second limitation of the present study lies in the fact that students were largely Hispanic, Native American, or mixed races. It is plausible that students with different cultural contexts may have different problem-solving outcomes.

Acknowledgments

Ideas in this manuscript stem from grant-funded research by the National Science Foundation (NSF #2100988, 2101026; #1720646, 1720661). Any opinions, findings, conclusions, or recommendations expressed by the authors do not necessarily reflect the views of the National Science Foundation. We also wish to acknowledge our gratitude to the students, schools, and project leadership associated with this research.

References

Barrow, M. A. (2014). Even math requires learning academic language. *Phi Delta Kappan*, *95*(6), 35-38.

Barwell, R., Moschkovich, J. N., & Phakeng, M. S. (2017). Language diversity and mathematics: Second language, bilingual, and multilingual learners. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 583–606). National Council of Teachers of Mathematics.

Bay-Williams, J. M., & Livers, S. (2009). Supporting math vocabulary acquisition. *Teaching Children Mathematics*, *16*(4), 238-246.

Bialik, K., Scheller, A., & Walker, K. (2018). *6 facts about English language learners in U.S. public schools*. http://tinyurl.com/4whvvabe

Bostic, J. (2021). Think alouds: Informing scholarship and broadening partnerships through assessment. *Applied Measurement in Education, 34*(1), 1-9. *DOI:* 10.1080/08957347.2020.1835914

Bostic, J., Sondergeld, T, Matney, G., Stone, G., & Hicks, T. (2021). Gathering response process data for a problem-solving measure through whole-class think alouds. *Applied Measurement in Education, 34*(1), 46-60*. DOI: 10.1080.08957347.2020.1835913*

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.

Bostic, J., Sondergeld, T., Folger, T. & Kruse, L. (2017). PSM7 and PSM8: Validating two problem-solving measures. *Journal of Applied Measurement, 18*(2), 151-162.

Bostic, J., Pape, S., & Jacobbe, T. (2016). Encouraging sixth-grade students' problem-solving performance by teaching through problem solving. *Investigations in Mathematics Learning, 8*(3), 30-58. https://doi.org/10.1080/24727466.2016.11790353

Bostic, J., & Sondergeld, T. (2015). Measuring sixth-grade students' problem solving: Validating an instrument addressing the mathematics Common Core. *School Science and Mathematics Journal, 115, 281-291*. **https://doi.org/10.1111/ssm.12130**

Buffington, P., Knight, T., & Tierny-Fife, P. (2017). Supporting mathematics discourse with sentence starters and sentence frames. *Maine: EDC.*

Bush, S. B., Karp, K. S., & Dougherty, B. J. (2020). *The math pact, middle school: Achieving instructional coherence within and across grades*. Corwin.

Creswell, J. (2011). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson.

de Araujo, Z., Roberts, S. A., Willey, C., & Zahner, W. (2018). English learners in K–12 mathematics education: A review of the literature. *Review of Educational Research*, *88*(6), 879-919.

Fuchs, L. S., Newman-Gonchar, R., Schumacher, R., Dougherty, B., Bucka, N., Karp, K. S., ... & Morgan, S. (2021). Assisting students struggling with mathematics: Intervention in the elementary grades (WWC 2021006). *National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, US Department of Education. http://whatworks.ed. gov*.

Gándara, P., & Hopkins, M. (2010). The changing linguistic landscape of the United States. In *Forbidden language: English learners and restrictive language policies* (pp. 10–17). essay, Teachers College Press.

Gorgorió, N., & Planas, N. (2001). Teaching mathematics in multilingual classrooms. *Educational studies in mathematics*, *47*(1), 7-33.

Guan, Z., Lee, S., Cuddihy, E., & Ramey, J. (2006, April). The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 1253-1262).

Jones, P. L. (1982). Learning mathematics in a second language: A problem with more and less. *Educational Studies in Mathematics*, *13*(3), 269-287.

Kena, G., Hussar W., McFarland J., de Brey C., Musu-Gillette, L., Wang, X., Zhang, J., Rathbun, A., Wilkinson-Flicker, S., Diliberti M., Barmer, A., Bullock Mann, F., and Dunlop Velez, E. (2016). *The Condition of Education 2016* (NCES 2016-144). U.S. Department of Education, National Center for Education Statistics. Washington, DC. http://nces.ed.gov/pubsearch.

Leighton, J. P. (2017). Using think-aloud interviews and cognitive labs in educational research. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199372904.001.0001

Lesh, R., & Zawojewski, J. (2007). Problem solving and modeling. In F. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 763-804). Charlotte, NC: Information Age Publishing.

Martin, D. B., & Larnell, G.V. (2013). Urban mathematics education. In H. R. Milner IV & K. Lomotey (Eds.), *Handbook of urban education* (pp. 373–393). Routledge.

Martiniello, M. (2008). Language and the performance of English-language learners in math word problems. *Harvard Educational Review*, *78*(2), 333-368.

Matney, G., Bostic, J., Fox, M., Hicks, T., May, T., & Stone, G. (2022). Fourth-grade students' sense-making during multi-step problem solving. *Journal for Mathematical Behavior*, 65, 1-17. *https://doi.org/10.1016/j.jmathb.2022.100933*

McGraw, R., & Rubinstein-Avila, E. (2009). Middle school immigrant students developing mathematical reasoning in Spanish and English. *Bilingual Research Journal*, *31*(1-2), 147-173.

Miles, M., Huberman, A.M., & Saldaña, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (3rd Ed.). Thousand Oaks, CA: Sage.

Moschkovich, J. N. (2010). Language(s) and learning mathematics: Resources, challenges, and issues for research. In J. N. Moschkovich (Ed.), *Language and mathematics education: Multiple perspectives and directions for research* (pp. 1–28). essay, Information Age Pub.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

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

National Education Association. (2020). *English language learners*. NEA. Retrieved from https://www.nea.org/resource-library/english-language-learners

Nevárez-La Torre, A. A. (2012). Transiency in urban schools: Challenges and opportunities in educating ELLs with a migrant background. *Education and Urban Society*, *44*(1), 3-34.

Ní Ríordáin, M., & O'Donoghue, J. (2009). The relationship between performance on mathematical word problems and language proficiency for students learning through the medium of Irish. *Educational Studies in Mathematics*, *71*(1), 43-64.

Pape, S. J. (2004). Middle school children's problem-solving behavior: A cognitive analysis from a reading comprehension perspective. *Journal for research in Mathematics Education*, *35*(3), 187-219.

Planas, N., & Civil, M. (2013). Language-as-resource and language-as-political: Tensions in the bilingual mathematics classroom. *Mathematics Education Research Journal, 25*(3), 361-378.

Rodriguez, D., Carrasquillo, A., Garcia, E., & Howitt, D. (2020). Factors that challenge English learners and increase their dropout rates: Recommendations from the field. *International Journal of Bilingual Education and Bilingualism*, *25*(3), 878-894.

Sanford, A. K., Pinkney, C. J., Brown, J. E., Elliott, C. G., Rotert, E. N., & Sennott, S. C. (2020). Culturally and linguistically responsive mathematics instruction for English learners in multi-tiered support systems: PLUSS enhancements. *Learning Disability Quarterly*, *43*(2), 101-114.

Stinson, D. W. (2006). African American male adolescents, schooling (and mathematics): Deficiency, rejection, and achievement. *Review of Educational research*, *76*(4), 477-506.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research* (2nd ed.). Newbury Park, CA: Sage.

Stufft, D. L., & Brogadir, R. (2011). Urban principals' facilitation of English language learning in public schools. *Education and Urban Society*, *43*(5), 560-575.

Sorto, M. A. (2019, October). Components of High-quality Mathematics Classrooms: Attending to Learning Opportunities for English Language Learners. In S. Otten, A. Candela, Z de Araujo, C. Haines, & C. Munter (Eds.), *Proceedings of the 41st Annual Meeting of PME-NA (pp. 1594-1603)*. University of Missouri, Columbia.

Thompson, D. R., & Rubenstein, R. N. (2000). Learning mathematics vocabulary: Potential pitfalls and instructional strategies. *The Mathematics Teacher*, *93*(7), 568-574.
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(3), 195–229. https://doi.org/10.1207/s15327833mtl0103_2
Verschaffel, L., Greer, B., & DeCorte, E. (2000). Making sense of word problems. Lisse, Netherlands: Swets & Zeitlinger.
Wells, A. (2014). *Seeing past the "colorblind" myth of education policy: Addressing racial and ethnic inequality and supporting culturally diverse schools*. Boulder, CO: National Education Policy Center.
Welsh, R. & Swain, W. (2020). (Re)defining urban education: A conceptual review and empirical exploration of the definition of urban education. *Educational Researcher, 49*(2), 90-100.
WIDA. (2020). WIDA *English language development standards framework, 2020 edition: Kindergarten – grade 12.* Board of Regents of the University of Wisconsin System.
Yee, S., & Bostic, J. (2014). Developing a contextualization of students' mathematical problem-solving. *Journal for Mathematical Behavior*, *36,* 1-19.
