Relations Between Mathematical Vocabulary and Children's Mathematical Performance

by

Shujie Charlene Song

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Arts

In

Psychology

Carleton University
Ottawa, Ontario

Abstract

Mathematical vocabulary comprises terms that have mathematical meanings and which people use to communicate mathematical concepts (e.g., large, digit, circle, equal sign). Students' general vocabulary and their mathematical vocabulary are both correlated with mathematical skills. I asked whether mathematical vocabulary fully or partially mediates the relation between general vocabulary and mathematical performance. Canadian students in grade 3 ($N = 234$, mean age = 8.7 years) completed measures of general and mathematical vocabulary and several different mathematical outcomes (i.e., arithmetic fluency, pre-algebra, and word-problem solving). Students were either learning mathematics in English or in French; the latter group were in immersion schools and thus had English as their first language. The results showed that students' mathematical vocabulary partially mediated the relation between general vocabulary and applied mathematical skills (i.e., pre-algebra and problem solving), and fully mediated the relation between general vocabulary and arithmetic fluency; and second, that students' domain-general cognitive skills (e.g., working memory, nonverbal reasoning) partially mediated the relation between mathematical vocabulary and applied mathematical skills, but not arithmetic fluency. Lastly, Numeration Words partially mediated the relation between general vocabulary and applied mathematical skills (i.e., pre-algebra and problem-solving), and fully mediated the relation between general vocabulary and arithmetic fluency. These analyses provided information about how individual differences in domain-specific and domain-general skills are related to students' mathematical performance.

Acknowledgements

I would like to take this opportunity to thank everyone who helped and supported me in completing my thesis. First, I would like to thank Dr. Jo-Anne LeFevre, my supervisor, who provided me with ongoing guidance, support, care, patience, and excellent editing skills. I truly learned a lot from this process. I am very proud and honored to having worked with Dr. Jo-Anne LeFevre.

Thanks to my thesis review committee of Dr. Monique Senechal and Dr. Beth MacLeod who offered constructive feedback and thoughtful questions. Thanks also to Dr. Chang Xu in the math lab who have patiently advised me on statistical methods and analysis. I would like to thank the LLAMA team as well, for organizing the LLAMA data and sharing the data with me, also for their continuous support.

And finally, thanks to my family. My family was my greatest motivation in completing my master's degree. They supported me going back to school to do what I truly want to do, they are my biggest support both in my work and at home. Especially my husband, Victor, thanks for your love, patience, and support.

Table of Contents

List of Tables

List of Figures

List of Appendices

Introduction

Mathematics provides an essential foundation for students' success in disciplines such as science, technology, and engineering (Baker & Galanti, 2017; English, 2017). However, mathematics, including mathematical language, is challenging for many students (Riccomini et al., 2015; Schleppegrell, 2007). Mathematics involves many different cognitive skills, including language (Peng et al., 2020), working memory (Raghubar et al., 2010), and nonverbal reasoning (Morsanyi et al., 2018). Thus, it is important to understand how language and other cognitive skills are related to mathematical learning. Findings from a meta-analysis (Peng et al., 2020) suggested that language is an important medium used by students to communicate, understand, and retrieve mathematical concepts. Both general language skills and knowledge of math-specific language are related to mathematics learning (Peng et al., 2020; Lin et al., 2021).

One aspect of mathematical language is vocabulary, which refers to terms that are either unique to mathematical contexts, such as 'equal sign', 'denominator' or 'sphere', or terms used in a specific way in mathematical contexts, such as 'product' or 'sum' (Harmon et al., 2015). Mastery of mathematical vocabularies that are tied to mathematical context allows students to communicate mathematical concepts, and can help them to construct mathematical meanings, build representations of mathematical concepts, and retrieve information from long-term memory while they are performing mathematical tasks (Peng et al., 2020). The goals of the present research were to examine whether mathematical vocabulary fully or partially mediates the relation between general vocabulary and mathematical performances (i.e., arithmetic fluency, pre-algebra, and word-problem solving) for students in grade 3 and to determine whether the subcategories of mathematical vocabulary (i.e., Numeration and Geometry-spatial categories) are differentially related to mathematical outcomes.

Mathematical Vocabulary and Mathematics Development

Mathematical vocabulary encompasses the terms that people use in mathematical communication, reasoning, and computation (Brown, 1994), especially during the pedagogical conversation between students and teachers (Riccomini et al., 2015). The relation between mathematical vocabulary and mathematical performance has been studied by a number of researchers in the past decade (Bae et al., 2015; Fuchs et al., 2015; Kostos & Shin, 2010; Powell et al., 2017; Purpura & Logan, 2015; Purpura & Reid, 2016; Toll & Van Luit, 2014). Researchers examined this relation using children with different age groups and found correlations between mathematical vocabulary and different mathematics performances were moderate and significant (Fuchs et al., 2015; Powell et al., 2017; Purpura & Logan, 2015; Purpura & Reid, 2016; Toll & Van Luit, 2014; Ünal et al., 2021). In a meta-analysis study, Lin et al., (2021) summarized that the average correlation between mathematical vocabulary and mathematics performance was .49.

Beside the correlational studies, the results of a few intervention studies showed that increases in mathematical vocabulary skills: a) positively influenced grade 2 children's mathematical communications (Kostos & Shin, 2010); (b) increased word problem solving abilities for students in grades 5 (Amen, 2006); and (c) improved the performance of junior high school students on mathematics tests (Fletcher & Santoli, 2003).

In summary, there are strong correlations between mathematical language and mathematical performance and some evidence that training mathematical vocabulary supports students' mathematical development more generally.

Preschool Learners

Mathematical vocabulary is positively related to children's early numeracy performance. For preschool children, mathematical vocabulary includes terms such as more, less, near, and far. Purpura and Logan (2015) found mathematical vocabulary predicted early numeracy performance (i.e., number comparison, number order, number combinations) from fall to spring of the academic year for 114 children (M_{age} = 4.18 years). Purpura and Reid (2016) found that mathematical vocabulary, but not general vocabulary, significantly predicted children's numeracy performance when both measures were included in the model (N = 136, M_{age}= 4.3 years). Similarly, Toll and Van Luit (2014) showed the relation between children's early numerary and general language was mediated by children's mathematical language (N = 1030, M_{age} = 4.56 years). In summary, the findings of these studies showed that mathematical vocabulary plays an important role in children's acquisition of the early mathematical knowledge as early as preschool.

Elementary School and High School Learners

Mathematical vocabulary is also related to mathematical performance for older children. Once children enter school, they are expected to learn the meanings of mathematical terms such as addition, decimal, and hundreds. Fuchs et al. (2015) assessed 206 American children (M_{age} = 7.6 years, grade 2) on general vocabulary, working memory, basic arithmetic fluency, and math-specific vocabulary, and tested the relations between these measures and math word-problem solving. Mathematical vocabulary was measured by asking students to produce a word or recall the meaning of a word. The mathematical vocabulary measure included comparative, relational, and quantitative words such as bigger, smaller, more, and fewer. Fuchs et al. found that the relation between general vocabulary and word-problem solving was mediated by math-specific vocabulary. Furthermore, Powell et al. (2017) found relations among general vocabulary, mathematics computation, and mathematical vocabulary performance for students in third grade (M_{age} = 9.5 years) and fifth grade (M_{age} = 11.4 years). The results showed that general vocabulary and mathematics computation were significant predictors of mathematical vocabulary, and the relations among general vocabulary, mathematics computation, and mathematical vocabulary were stronger in third grade than in fifth grade. Powell et al. (2017) concluded that younger students are more likely to rely on general vocabulary to learn mathematics and mathematical vocabulary whereas older students are more likely to rely on mathematical vocabulary to learn mathematics.

Ünal et al. (2021) extended the work on mathematical vocabulary to older students from the US (N = 89, M_{age} = 14.2 years) and Turkey (N = 188, M_{age} = 13.9 years). Children were divided into higher- and lower achieving groups based on their performance on a mathematics computation test. Ünal et al. found that mathematical vocabulary was correlated with mathematics achievement. However, mathematical vocabulary only mediated the relation between general vocabulary and mathematics achievement for higher-achieving students. For the lower-achieving students, only general vocabulary predicted mathematics achievement. Ünal et al. (2021) argued that the finding showed the development of mathematical vocabulary may contribute to further learning in mathematics, which is consistent with the general conclusion of Powell et al. (2017).

In summary, the results of these studies (Fuchs et al., 2015; Powell et al., 2017; Ünal et al., 2021) extended findings from studies of preschool children and showed that mathematical vocabulary is related to early numerical knowledge and also facilitated the learning of advanced mathematical skills. Given the significant role of mathematical vocabulary in learning mathematics, it is important for educators to understand how mathematical vocabulary and mathematical performance are linked.

Models of the Relations Between Language and Mathematics

Harmon et al. (2005) argued that mathematics has its own language, which is complex, abstract, and different from general language, and describes mathematical content (e.g., number, geometry, operations, shapes, space, and so on). On this view, mathematical vocabulary is a central feature of students' developing mathematical knowledge. Peng et al. (2020) used the *Function Hypothesis of Language* to further explain the relations among general language, mathematical language, and mathematics learning. The *Function Hypothesis of Language*, "… suggests that language serves many functions in our lives, such as exchanging/delivering messages/information, expressing our feelings and attitudes, and informing our thoughts (Bruner, 1966; Fetzer & Tiedemann, 2018; Vygotsky, 1986)" (Peng et al., 2021; p. 596). Peng et al. further distinguished between two specific functions of language. The *medium function* hypothesis is that language is used as the medium for people to connect their mathematical learning with that of others. This function allows people to grasp the meanings of mathematical information, either through use of general words in mathematical contexts (e.g., less, more), or use of specific words that are encountered only in mathematical contexts (e.g., denominator, subtraction). According to the *medium function hypothesis*, mathematical concepts are stored in a language format which has specific vocabulary, which implies that students' knowledge of mathematical language will be a better predictor of mathematical performance skills than their general language skills (O'Halloran, 2005; Peng & Lin, 2019, see Figure 1).

The second specific function of language, the *thinking function,* also identifies language as a tool for learning and performing mathematics. However, the *thinking function* emphasizes the involvement of other cognitive abilities (e.g., working memory and intelligence) which people use to "think about more abstract mathematical concepts and relations between them [i.e., between the concepts]" (Peng et al., 2020; p. 597). Peng et al. hypothesized that working memory and nonverbal reasoning are domain-general cognitive skills that may confound the relations between mathematical language and mathematics, and they can explain part of the relation between mathematical language and mathematics (see Figure 2). Specifically, these cognitive abilities influence how people use mathematical vocabulary, and therefore, the use of mathematical vocabulary affects people's ideas about how to understand and organize mathematical concepts.

Figure 1. *Medium Function Hypothesis*

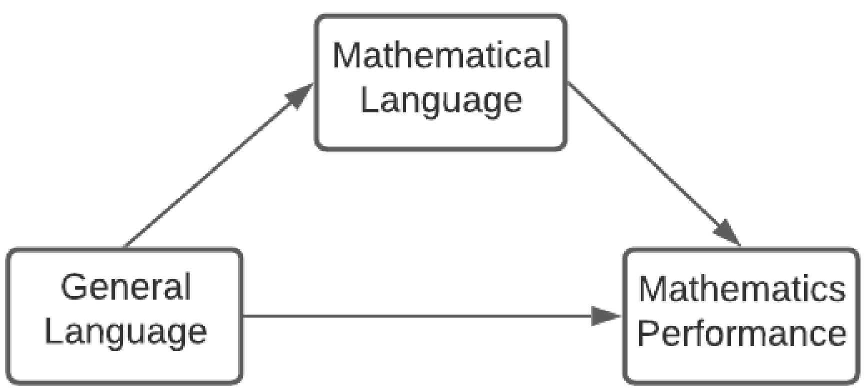
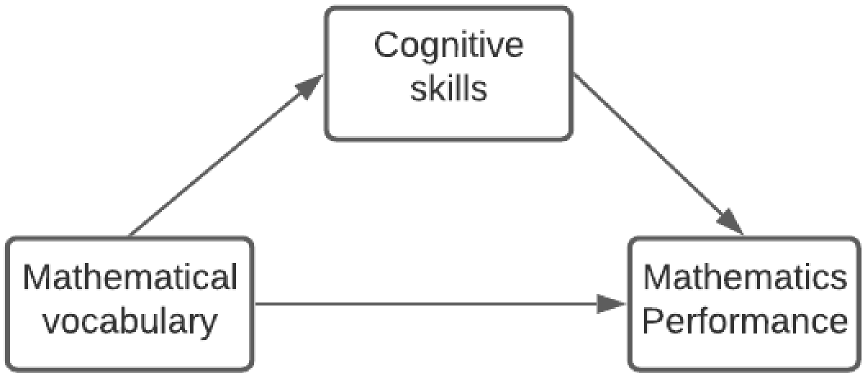


Figure 2. *Thinking Function Hypothesis*



The *Function Hypothesis of Language for Mathematics* (both *medium function* and *thinking function*) (Peng et al., 2020) was supported by another meta-analysis (Lin et al., 2021) which found an average correlation of .49 between mathematical vocabulary and mathematics performance, and the correlation remained moderate and significant after controlling for both language comprehension (i.e., reading comprehension, listening comprehension, and general vocabulary) and cognitive skills. Lin et al. also found a trend such that the relations between higher-order mathematical tasks (e.g., word-problem solving) and mathematical vocabulary are stronger than the relations between fundamental mathematical tasks and mathematical vocabulary (e.g., number operation, arithmetic fluency). Lin et al. suggested that mathematical vocabulary is not just a proxy for students' mathematical skills, but that they actively use mathematical vocabulary as a medium to grasp the meanings of mathematical information and to retrieve mathematical knowledge to perform mathematical tasks. Therefore, a close investigation of mathematical vocabulary is important for educators to understand the mechanism of mathematics learning in children.

Measuring Mathematical Vocabulary

Mathematical vocabulary was treated as a unitary construct in many studies (Fletcher & Santoli, 2003; Kostos & Shin, 2010; Purpura, Logan, et al., 2017; Purpura & Reid, 2016; Ünal et al., 2021). Other researchers, however, have suggested that math vocabulary is not a unitary construct (Fuchs et al., 2015; Monroe & Panchyshyn, 1995; Peng & Lin, 2019; Powell & Nelson, 2017). In this section, I discuss several different ways in which mathematical vocabulary has been described. I note, however, that in most empirical work with school aged children, vocabulary terms have been selected to represent curriculum areas (e.g., numeracy, geometry), with little attention to other characteristics of the terms or their inter-relations. For preschool children, mathematical vocabulary tends to overlap with words that are also used in everyday life (e.g., larger, smaller, behind, after; Monroe & Panchyshyn, 1995; Purpura & Logan, 2015).

Curriculum/Area Based Categorizations

One source of variability in the measurement of mathematical vocabulary is whether the assessment only includes mathematical terms, or whether the mathematical terms are a subset of items from an existing measure. For example, Toll and Van Luit (2014) measured mathematical vocabulary using a subset of 22 math-related words that were part of an existing measure of general vocabulary. The limitation of this approach is that there was no systematic approach to item selection. In contrast, most researchers created new measures of mathematical vocabulary where the items were selected from curricula resources (Douglas, 2020; Fuchs et al., 2012; Purpura & Logan, 2015). For example, Purpura and Logan (2015) designed an assessment that was based on items from preschool curriculum materials. The Preschool Assessment of the Language of Mathematics (PALM) contains 16 items and is closely related to curriculum requirements. Thus, it may better capture key vocabulary terms for preschool children's mathematical learning. Because of the preschool focus, however, the measure is not suitable for older children. Measures designed for older children additionally include more technical mathematical vocabulary, as described in the following session.

Another source of variability in the measurement of mathematical vocabulary is whether the measurement was treated as unitary construct or not. In most empirical studies of mathematical vocabulary for school-aged children, researchers categorized words according to the grade-level and/or the content-specific domains in mathematics curricula (e.g., Common Core States Standards 2010; Ontario Ministry of Education, 2020). First, mathematical vocabulary can be categorized by grade level of introduction based on the curriculum. Forsyth and Powell (2017) identified 135 terms presented in Common Core State Standards (2010), and then coded the terms by the grade level at which a mathematical term was first introduced, from kindergarten to sixth grade. They assumed that students should be more familiar with the items introduced at lower grades than the items introduced at the higher grades. In addition, categories of mathematics vocabulary were identified by the curriculum designers to help teachers focus on a certain domain of mathematical concepts (i.e., geometry, arithmetic fluency, and so on) as they teach students. Forsyth and Powell (2017) identified four domains for both grade 1 and grade 2 in the Common Core States Standards (2010): 1) operations and algebraic thinking; 2) number and operations in base ten; 3) measurement and data; and 4) geometry. Fractions was added as the fifth category for students in grades 3 to 5. The same measure was used by Powell et al. (2017) and they treated it as an unitary construct, also Powell and Nelson (2017) selected 64 first-grade mathematical terms from this measure, but they categorized items in different ways in their study (more details will be described in the following section).

In contrast, Purpura and colleagues defined two content-specific categories of mathematical vocabulary that are broader than the curriculum categories: quantitative terms (e.g., most, more, a little bit, and less) and spatial terms (e.g., nearest, below, last, and before) (Purpura et al., 2017; Purpura & Reid, 2016). However, they treated mathematics vocabulary as a unitary construct in subsequent analyses. Purpura and colleagues (2016; 2017) recognized that the analysis of mathematical vocabulary in their studies were restricted, and they suggested that future studies should examine the relations among different domains of mathematical vocabulary and mathematical skills.

Complexity/Familiarity Based Categorizations of Mathematical Vocabulary

Monroe and Panchyshyn (1995) described four categories of mathematical vocabulary that are based on the specificity and complexity of the words in relation to their mathematical meanings. These categories include: 1) *technical terms* that have one meaning that represents mathematical concepts (e.g., integer); 2) *sub-technical terms* that have multiple meanings and represent both mathematical concept and daily experience (e.g., volume; degree); 3) *symbolic* terms that are Arabic numbers and symbols represent mathematical concepts (e.g., numerals "1"," 2", numerical expression "4^2" four-squared and "+"); and 4) *general vocabulary* that are terms commonly encounter in mathematics and everyday language (e.g., take away). Monroe and Panchyshyn (1995) argued that, because learning mathematical vocabulary is difficult for students due to the complexity and abstraction of the words, categorization of mathematical vocabulary would help teachers to develop different teaching strategies to actively teach these four kinds of vocabulary. Most of the terms used in mathematical vocabulary studies with children between the ages of four and six fall into the last category of everyday language (e.g., a little bit, under, or take away, Purpura & Logan, 2015). The more specific and technical terms, and the symbolic terms, are usually introduced in school (as shown in Powell & Nelson, 2017).

Rubenstein and Thompson (2002) similarly categorized mathematics vocabulary based on the meanings of the terms, but they also considered the difficulty of learning and comprehension. They proposed 11 categories of mathematical language: 1) terms used in mathematics and everyday situations with more than one meaning (e.g., right, foot); 2) terms used in mathematics and everyday situations, with a more precise definition in mathematics (e.g., odd); 3) terms that are used only in mathematics and have one mathematical meaning (e.g., subtract); 4) terms used only in mathematics that have more than one mathematical meanings (e.g., round as a circle vs. to round a number to the tenth place); 5) terms shared in both mathematics and other disciplines with different definitions; 6) terms appear in mathematics context that have homophones or homographs (e.g., variable in mathematics represents possible numerical values, but variable clouds in science are a weather condition); 7) terms that appear in mathematical contexts that have related but distinct meanings (e.g., numerator and denominator); 8) a single English mathematical term that may have many different meanings in other languages (e.g., an ordinary table in Spanish is *mesa*, but a mathematical table is *tabla* in Spanish); 9) terms with spelling irregularities (e.g., fraction denominators, such as sixth, fifth, fourth, and third, are like ordinal numbers, but rather than second, the fraction is half) ; 10) terms that have different ways of expression (e.g., skip count by three vs. multiplies of 3); and 11) terms that are informal ways of describing mathematical concepts (e.g., diamond for rhombus) (Rubenstein & Thompson, 2002). In short, this categorization suggested that mathematical language is very complex and therefore it is difficult for students to understand. The authors concluded that teachers should be aware of the potential confusion students may experience, given the complexity, and educators should teach mathematical vocabulary adequately. However, these authors did not empirically determine whether their categorization would be useful for teachers.

Summary

In summary, researchers have devised various measures of mathematical vocabulary. The nature and content of these measures varies according to the age of the participants. For preschool students, the focus is on terms that are also often used in everyday life (such as “more” and “take away”, Purpura & Logan, 2015) whereas for older students, knowledge categories such as geometry may be used to select and group vocabulary items (such as "rotation" and "dilation", Ünal et al., 2021). Mathematical vocabulary terms have also been categorized according to complexity. Only some of these categorizations have been used in empirical studies, however, as described below.

Empirical Studies on Categories of Mathematical Vocabulary

In two studies, researchers examined the differences in children’s knowledge of categories of mathematical vocabulary (Forsyth & Powell, 2017; Powell & Nelson, 2017), and in one study, they examined the relations among mathematical vocabulary categories and mathematical performance (Peng & Lin, 2019). Powell and Nelson (2017) used three different ways of categorizing mathematical vocabulary and examined the differences in accuracy of 104 grade 1 American students’ responses on mathematics vocabulary measures (M_{age} = 7.4 years). The mathematics vocabulary measure included 64 items. First, they grouped 64 items into four content-specific domains (i.e., Operations and Algebraic Thinking, Number and Operations in Base 10, Geometry, and Measurement and Data) as suggested by Common Core States Standards (CCSS, 2010). Second, Powell and Nelson organized the mathematics vocabulary based on the four categories (i.e., technical, subtechnical, symbolic, general) proposed by

Monroe and Panchyshyn (1995). Third, they divided mathematics vocabulary into the 11 categories described by Rubenstein and Thompson (2002). They evaluated students' response accuracy among different domains and across these categorization approaches.

Powell and Nelson (2017) found significant differences in students' performance when terms were organized by CCSS domains. Specifically, students had more accurate answers on terms associated with the Measurement and Data than any other CCSS domains and they had more accurate answers on Geometry terms than on Number and Operations in Base 10 or Operations and Algebraic Thinking. Powell and Nelson also found significant differences among the four technical categories: General terms were easiest for students (91.1% correct), followed by subtechnical (56.4%), symbolic (54.5%), and technical (42%). In contrast, there were no significant differences in accuracy among the 11 categories proposed by Rubenstein and Thompson. In short, Powell and Nelson showed that there were significant differences in students' response accuracy between vocabulary categories based on CCSS and on technical complexity/familiarity from use in everyday life.

Forsyth and Powell (2017) further examined whether the differences among CCSS domains and grade levels were related to students' learning difficulties using the measure developed by Powell et al. (2017). They compared typically developing students, students with mathematical difficulties but no other learning difficulties (MD-only), students with reading difficulties only (RD-only), and students with both MD and RD (MDRD). Participants were selected from seven fifth-grade classrooms ($N = 128$, $M_{age} = 11.4$ years). They completed a general vocabulary measure, a math computation proficiency measure, and a mathematical vocabulary measure. Although Forsyth and Powell used the same mathematical vocabulary measure used in Powell et al. (2017), instead of treating the measure as a unitary construct, they categorized mathematical vocabulary into four CCSS groups (i.e., Whole number, Fractions, Measurement, and Geometry).

Forsyth and Powell (2017) first examined whether there were differences among typically developing students and those in the difficulty groups (i.e., MD-only, RD-only, and MDRD) in mathematical vocabulary based on CCSS domains and grade level of term inclusion in textbook glossaries. They found performance of mathematical vocabulary did not vary by CCSS categories within any of the groups, which suggests that specific CCSS domains have equally difficult vocabulary. Forsyth and Powell did find differences among four groups of students based on grade levels of the mathematical terms, however. For terms from kindergarten through fourth grade, typical learners did significantly better than other three groups, there was no difference between MD-only and RD-only groups, and MDRD had the lowest performance. However, for terms introduced at fifth grade, MD-only and RD-only students performed as poorly those with MDRD.

Additionally, Forsyth and Powell (2017) examined students' responses on individual items to determine which mathematical vocabulary causes the most difficulty. They found that the overall accuracy rate was less than 75% for terms introduced at fourth grade and later. Based on their results, Forsyth and Powell suggested that mathematical vocabulary development should be addressed in the early grades. The level of mathematical vocabulary difficulty experienced should be the same across the mathematics curricula. Additionally, more exposures of mathematical vocabulary may help students master new terms.

Peng and Lin (2019) measured mathematical performance of Chinese students in grade 4 ($N = 237$; $M_{age} = 10.2$ years) on calculation, math word problems, general vocabulary, and mathematical vocabulary. In multiple regression analyses, they found that mathematical vocabulary made a unique contribution to math word problem solving but not to calculation after controlling for general vocabulary and other cognitive measures, which is consistent with the *Function Hypothesis of Language for Mathematics* (Peng et al., 2020). Furthermore, they divided mathematics vocabulary into three categories (i.e., Measurement, Geometry, and Numerical Operations) based on Chinese mathematics curriculum in elementary school. Measurement and Geometry vocabulary (a) partially mediated the relation between word problem solving and general vocabulary, (b) partially mediated the relation between word problem solving and IQ, and (c) fully mediated the relation between word problem solving and working memory (Peng & Lin, 2019). In contrast, Numerical Operations knowledge was correlated with both word problem solving and calculation but did not account for unique variance in any analyses. Peng and Lin explained that the Numerical Operation category did not contribute to word problem solving because it did not uniquely contribute to calculation (the foundation for word problem solving) and also the calculation was controlled in the analyses.

In summary, results of three studies supported the hypothesis that mathematical vocabulary is a multi-dimensional construct. Both Peng and Lin (2019) and Powell and Nelson (2017) found that the difficulty level varied among CCSS categories, and that the correlations between mathematical vocabulary and mathematical performances varied depending on the CCSS categories of the vocabulary and different mathematics skills. Forsyth and Powell (2017) suggested the difficulty level of mathematical vocabulary is different based on the grade level of that vocabulary was introduced. In other words, an item introduced in grade 5 is more difficult to master than an item introduced in grade 1. These researchers suggested categorization of mathematical vocabulary should be considered in future studies, on the assumption that acquisition of mathematical vocabulary will facilitate children's understanding of mathematics

(Hughes et al., 2020; Lin et al., 2021; Monroe & Panchyshyn, 1995; Rubenstein & Thompson, 2002).

Methodologies for Assessing Mathematical Vocabulary

Mathematical vocabulary can be measured using different types of memory demands (i.e., receptive/recognition vs. expressive/recall) and different formats (i.e., short answers, multiple choice). With receptive questions, terms are given to students (in oral or written form), and students are asked to choose the right pictures that corresponds to the given terms. For example, students were asked to select a picture that showed a “denominator”. With expressive questions, students may be asked to produce a term according to a specific description (e.g., “Draw a right angle” or “What is rotation”) or they may be asked to write a definition of the term. The requirement to recall either the term or the definition are much more difficult and also harder to administer and score (Peng et al., 2020).

Powell and colleagues developed several mathematical vocabulary assessments that were described in the previous section (Forsyth & Powell, 2017; Powell et al., 2017; Powell & Nelson, 2017). In this section, the focus is the methodology of each assessment. Powell and Nelson (2017) included 64 mathematical terms in the grade 1 assessment and included recall questions, comprehension questions, and application questions. Similarly, the assessments developed by Powell et al. (2017) for grades 3 and 5 contain 133 math-related words administered with multiple question types, including multiple choice, word and definition matching, and short answers. In short, Powell and colleagues selected both expressive questions (e.g., comprehension and short answer) and receptive questions (e.g., multiple choices) in their mathematical vocabulary assessments.

More recently, this research group (i.e., Hughes et al., 2020) developed a math vocabulary assessment for students in grades 7 and 8 that included 57 terms and a single type of question (i.e., receptive measure with multiple choice). Similarly, Douglas (2020) developed a math vocabulary assessment using receptive items only and tested it with students in grades 4 and 6. This measure was administered on an iPad, and students were asked to choose, from four pictures, which is the best match of a math word or phrase. Douglas' measure is very similar to standardized tests of general vocabulary and involves only receptive vocabulary knowledge. According to Peng et al. (2020), measuring language in either an expressive format or a receptive format would not affect the accuracy of those measures. In the present research, Douglas's measure, which she called Math Words, was used.

Comparison of Mathematical Vocabulary Measures

According to Powell et al. (2021), by grade 6, students would have experienced hundreds of mathematical vocabulary words. Thus, researchers studying mathematical vocabulary are forced to choose a subset of the possible words that students might know, and so no single measure will be comprehensive. The Math Words assessment used in the present research was designed by Douglas (2020) for use with students in Ontario and Manitoba. Note that Douglas created her measure before Powell and colleagues had published their comprehensive paper on mathematical vocabulary (Forsyth & Powell, 2017; Powell et al., 2017). Douglas created a master list of over 200 words using curriculum documents and mathematics textbooks (Common Core State Standards Initiative, 2010; Manitoba Education, 2013; Ontario Ministry of Education, 2005). The master list was classified by grade and by the five mathematics strands described in the Ontario curriculum (i.e., Numeration, Measurement, Geometry-Spatial, Patterning and Algebra, Data management and Probability). Two versions of the Math Words measures were created, one for children in grade 2 to grade 3 (i.e., Version I), and a revised version for children in grade 3 to grade 6 (i.e., Version II). Version II was used in this current study. It was created by narrowing down the master list of 200 words to 40 items by selecting key words from five mathematics strands with a heavier emphasis on numeration words. Numeration words were emphasized because they were considered as fundamental items which are used more frequently than other categories (Douglas, 2020).

The methodology of the Math Words measure was closely modelled on a measure of general receptive vocabulary, the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997). Thus, each word was represented by a picture and three other pictures were selected as foils. Math Words is therefore a receptive vocabulary test administered in multiple-choice format. One possible strength of this measure is that it includes terms from grades 1 to 8 to avoid floor or ceiling effects (Douglas, 2020). Math Words includes items from four curriculum strands: Numeration (n = 23; Table 1), Geometry-Spatial (n = 10; Table 2), Measurement (n = 5; Table 3), and Data (n = 2; Table 4). To explore how similar this measure is to those used in other studies, these four sets of words were compared to those used in the three other papers that used curriculum-based categories (Forsyth & Powell, 2017; Peng & Lin, 2019; Powell & Nelson, 2017).

As shown in Table 1, 23 items from Math Words were selected from the Numeration category. Only four of these items were used by Powell and Nelson (2017) but their study was with grade 1 students. In contrast, 18 of the Math Words Numeration items overlap with items used by Forsyth and Powell (2017). Note that these authors had fraction-related items (i.e., unit fraction, numerator, and denominator) in a separate category from *Numeration*. Finally, eight of these items overlap with those from the *Numerical Operation* category from Peng and Lin

(2019), which included words for grades 3, 4, and 5. In summary, all but four of the words in the *Numeration* set from Math Words were also used in other studies, indicating considerable conceptual similarity across the measures, especially with those designed for students in grades 3, 4, and 5.

Table 1. *Numeration Terms (n =23)*

#	Douglas (2020): Grade 4 and 6 *N* = 23	Grade[a]	Powell & Nelson (2017): Grade 1 *N* = 30	Forsyth & Powell (2017): Grade 5 *N* = 67	Peng & Lin (2019): Grade 4 *N* = 34
2	tens column	2	✓[1]	✓[1]	
5	hundreds column	2	✓[1]	✓[1]	
6	less than	2			
9	ascending	2			
8	sum	3	✓	✓	
10	difference	3	✓	✓	
11	array	3		✓	
13	product	3		✓	
22	unit fraction	4		✓[5]	
23	decimal	4		✓	✓[2]
24	hundredths column	4		✓[3]	✓[3]
18	factor	4			✓
20	denominator	4		✓[5]	✓
21	numerator	4		✓[5]	✓
27	equivalent fraction	5		✓[5]	
28	multiple	5		✓	✓
29	variable	5		✓	
26	quotient	6		✓	✓
36	prime number	6		✓[4]	✓
34	ratio	6			
35	negative number	7		✓[6]	
39	square root	7			
40	integer	7		✓[6]	

Notes. [a] This column indicates item was selected from which grade from Douglas (2020). [1] "Tens column" and "Hundreds column" is "Tens" and "Hundreds" in Powell and Nelson (2017) and Forsyth and Powell (2017). [2] "Decimal" is "Repeating decimal" in Peng and Lin (2019), [3] “Hundredths column” is “hundredths” in both Forsyth and Powell, and Peng and Lin. [4] “Prime number” is “Prime” in Forsyth and Powell. [5] Grouped into a “Fraction” category separate from Number and Operations. [6] Term was ‘negative integer’ or ‘positive integer’.

As shown in Table 2, Math Words included ten items from the Geometry-Spatial category. Both the Powell and Nelson (2017) and the Peng and Lin (2019) measures had more Geometry items but nevertheless, all but three items from Math Words were included in those other lists. In this category, the percentage of item overlap is 60% with Forsyth and Powell and 40% with Peng and Lin, although only 30% of items from Math Words appeared in both other studies. Selection of specific items clearly varies across studies, however, there does seem to be considerable conceptual similarity.

Table 2. *Geometry-Spatial Terms (n = 10)*

Item	Douglas (2020): grade 4 and 6	Grade[a]	Powell & Nelson (2017): grade 1	Forsyth & Powell (2017): grade 5	Peng & Lin (2019): grades 3, 4, 5
#	N = 10		N = 18	N = 47	N = 30
3	cylinder	1	✓	✓	
16	triangular prism	3	✓	✓	
12	horizontal	3			
31	clockwise	3			
25	parallel	5		✓[1]	✓[1]
30	acute angle	5		✓	✓
32	obtuse angle	5		✓	
33	equilateral triangle	6		✓	✓
37	perpendicular	7			✓
38	radius	8		✓	

Notes. [1] "Parallel" is "parallel line" in both Forsyth and Powell (2017) and Peng and Lin (2019). [a] Grade refers to the curriculum lists used by Douglas (2020) for Manitoba and Ontario.

The third grouping of terms used in Douglas (2020) is shown in Table 3 and was labelled "Measurement". Three of these items were also used by Forsyth and Powell. However, two items from Measurement that were included in Douglas (2020) were not used in the other studies: mass and axes. Note that time-related terms were not included in Douglas's measure but were included in the *Measurement and Data* category in Powell and Nelson (e.g., hour, minute), as well as in *Measurement* category in Peng and Lin (2019; e.g., year, month, day, hour, minute). In this category, the percentage of item overlap is 60% with Forsyth and Powell but only 20% with Peng and Lin, and only one of the items from Math Words appeared in both other studies. Selection of specific items clearly varies across studies and because there are relatively few items in this category, it is not clear that the conceptual overlap is strong with the other lists.

Table 3. *Measurement Terms (n = 5)*

Item	Douglas (2020): grade 4, 6	Grade[a]	Powell & Nelson (2017): Grade 1	Forsyth & Powell (2017): Grade 3, 5	Peng & Lin (2019): Grade 4
	N = 5		N = 17	N = 16	N = 28
1	dime	1	✓	✓	
4	one-quarter (area)	2	✓[1]	✓[1]	✓[1]
7	mass	3			
14	perimeter	3		✓	
15	axes	3			

Notes. [1] "One-quarter" is "Quarter" in the other three studies. [a] Grade refers to the curriculum lists used by Douglas (2020) for Manitoba and Ontario.

The *Data Management* category only contains two items as shown in Table 4, and the term "key" from Data Management domain that was included in the Douglas (2020) was not used in the other studies. The term "column" was used to refer to columns in an arithmetic problem in Powell and Nelson (2017). Thus, this category does not overlap conceptually with any of those used in previous studies due to limited items.

Table 4. *Data Management Terms (n = 2)*

Item	Douglas (2020): grade 4, 6	Grade[a]	Powell & Nelson (2017): Grade 1	Forsyth & Powell (2017): Grade 3, 5	Peng & Lin (2019): Grade 4
	N = 2		N = 17	N = 16	N = 28
17	column	4	✓[1]		
19	key	4			

Notes. [1] "Item categorized in "Number and Operations in Base 10" in Powell and Nelson (2017).
[a] Grade refers to the curriculum lists used by Douglas (2020) for Manitoba and Ontario.

For the Numeration and Geometry-Spatial categories, there was considerable overlap across the three lists. Some differences in the selection and categorization of items may be related to the different curriculum guidelines used to create the lists. Two measures were developed based on CCSS (2010) which is the American curriculum guideline (Forsyth & Powell, 2017; Powell & Nelson, 2017). Peng and Lin developed their list based on Chinese mathematical curricula. Mathematical terms were selected from wide range of grade levels by both Forsyth and Powell (i.e., kindergarten through grade 6), and Peng and Lin's (2019) (i.e., grade 3 through grade 5), but the list used in Powell and Nelson was created based on grade 1 mathematics curriculum only.

The Math Words list was also much shorter than those used in these other studies because it was part of a large battery of measures, rather than the sole focus of the research studies. In total, 70% of items from Math Words overlapped with those used in Forsyth and Powell, and 32.5% and 22.5% overlapped with those used by Peng and Lin and Powell and Nelson, respectively. Other words included by Peng and Lin, such as “estimation”, “distributive law of multiplication”, “least common multiple”, and “composite number” were included in the Number Operation category in the study conducted by Peng and Lin. However, these items were not listed under the five mathematics strands described in Ontario Curriculum, and therefore, they were not included in the Math Words measure developed by Douglas (2020). In short, although the mathematical vocabulary measures were developed at the different times and based on different curricula, but they have similarities. The terms included in Numeration and Geometry-Spatial in the Math Words measure seem to overlap conceptually with similar categories in Forsyth and Powell and Peng and Lin. Thus, analyses of the categories in the present study will be focussed on those two subcategories.

Research Using the Math Words Measure

Douglas (2020) developed the Math Words measure in two iterations. Version I had a slightly different set of items and did not consistently include three foils for each item (i.e., a few had only two items). Version I was used in Year 1 of a project on Language Learning and Math Achievement (LLAMA) and included in a publication (Xu et al., 2022). The second, revised version, was used in Year 2 of the LLAMA project (i.e., Version II) and by Douglas (2020). Although some papers have been published using Year 2 data from the LLAMA project, the Math Words measure is not included in them (Song et al., 2021; Xu et al., 2022). Thus, the analyses in this thesis were based on data that had not been previously published or analyzed.

Xu et al., (2022) used Version I of Math Words to examine the role of mathematical vocabulary in arithmetic fluency and word problem solving for grade 2 and 3 students learning math in either their first language (n = 103, M_{age} = 8.1) or a second language (n = 57, M_{age} = 8.2). Xu et al. found that mathematical vocabulary was related to arithmetic fluency for both groups. Moreover, mathematical vocabulary partially mediated the relation between general vocabulary (measured with a subset of items from the PPVT) and word problem solving performance for first-language learners whereas only general vocabulary predicted performance for second-language learners. These findings are consistent with the medium function hypothesis by showing mathematical vocabulary mediated the relation between general vocabulary and mathematical performance.

Douglas (2020) used the Version II of the Math Words measure with students in grades 4 and 6. She treated the measure as a unitary construct and found that mathematical vocabulary directly predicted sixth graders' fraction skills. Moreover, mathematical vocabulary directly predicted the growth in fraction mapping skills for fourth graders. These findings showed that mathematical vocabulary plays an important role in children's mathematical learning.

In both these studies, the Math Words measure (Versions I and II) were reliable and showed expected patterns of correlations with math performance and general vocabulary. Both versions of the Math Words contain items from different math domains and items from different grade levels, but they were treated as unitary measures. In the current study, I first treated the Math Words measure as a unitary construct, and then broke it down into domain-level measures. I analyzed data from two samples from the LLAMA Project, not previously published, in which Version II of the Math Words measure was used with students in grade 3.

Current Study

The data I used in the current study was part of the LLAMA Project, which focused on the role of language in mathematics learning. Students in the larger LLAMA study were from a variety of language contexts including English, French, and Irish instructional milieu. Children completed four types of measures, including general cognitive measures (Spatial Span, Black White Stroop, Digit Forward/Backward, Matrix Reasoning), general linguistic measures (WIAT, PPVT, Mathematical Vocabulary), math-related linguistic measures (Mathematical vocabulary, Symbol Decision Task), basic symbolic mathematics measures (Order Judgment, Number Comparison/Ordering), and mathematical performance measures (Number Line Estimation, Arithmetic Fluency, Word Problem Solving, Pre-Algebra, and Measurement). Note that Version II of Math Words was used in Year 2 to measure children's mathematical vocabulary. For the present study, I focussed on two samples of students, one from Ottawa and the other from Manitoba, who completed the same English mathematical vocabulary measures (Math Words) in grade 3.

Many studies have shown that mathematical vocabulary is positively correlated with mathematical performance on measures such as arithmetic fluency and word-problem solving ability (Bae et al., 2015; Fuchs et al., 2015; Purpura & Logan, 2015; Xu et al., 2022). Thus, in the present research, I tested the *Function Hypothesis of Language* for Mathematics (Peng et al., 2019) by examining the relations among cognitive skills (i.e., working memory and reasoning ability), general vocabulary, mathematical vocabulary, and mathematical performance measures (i.e., arithmetic fluency, word problem solving and pre-algebra). I formulated three hypotheses. For the first two hypotheses, I used a unitary measure of math vocabulary. I used all 40 items from Math Words, and examined the internal reliability of the measure before using it in further analyses.

Hypothesis 1. For these students in grade 3, mathematical vocabulary will mediate the relation between general vocabulary and mathematical skills. This prediction is based on the '*medium function hypothesis*' (Peng et al., 2020) in which mathematical vocabulary mediates the relation between general vocabulary and mathematical skills (see Figure 1).

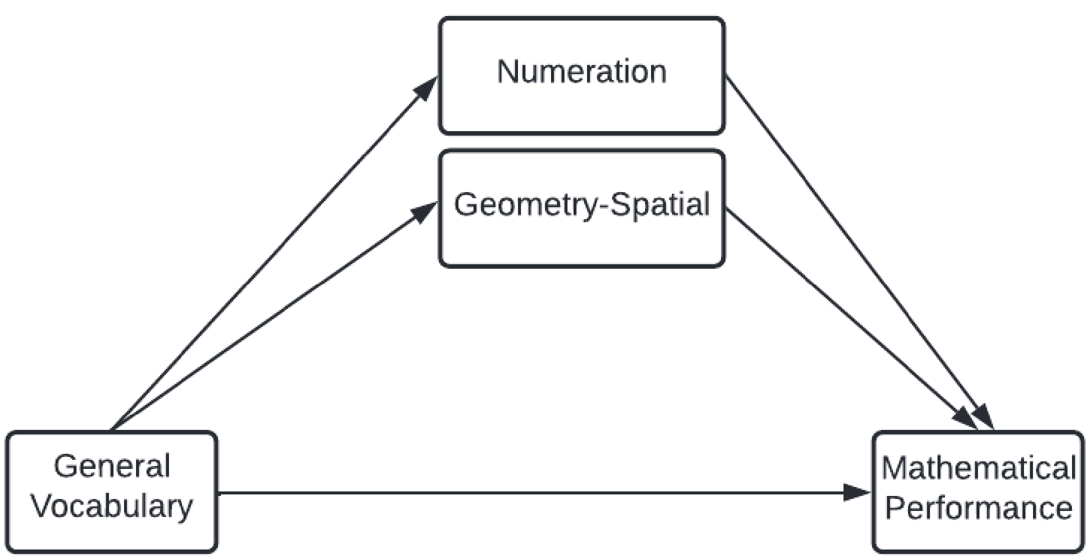
Hypothesis 2. Students' cognitive skills will mediate the relation between mathematical vocabulary and mathematical skills. This prediction is based on the '*thinking function hypothesis*' (Peng et al., 2020) in which children's cognitive abilities influence the relation between mathematical vocabulary and mathematical skills (see Figure 2).

The other goal of this thesis was to explore the view, first tested by Peng and Lin (2019) that the relations between mathematical vocabulary and performance are very specifically linked to content areas. Only a handful of studies examined different categories of mathematical vocabulary (Peng & Lin, 2019; Powell & Nelson, 2017), even though dividing mathematical vocabulary into content-specific domains is consistent with mathematics curricula in many countries (e.g., Common Core States Standards in US, Ontario Ministry of Education in Canada, and Chinese mathematics curriculum). Recall that Peng and Lin (2019) found that children's overall mathematical vocabulary skill accounted for unique variance in word problem solving. When they examined specific categories of math vocabulary knowledge, they found that Geometry and Measurement categories partially explained the relation between word problem solving and general vocabulary. Numerical Operations vocabulary did not account for unique variance in any analyses. To test whether such specificity is found in the present study, the Math

Word measure was broken down into four content-specific categories, and only Numeration and Geometry-Spatial subcategories were used to test Hypothesis 3.

Hypothesis 3. Different types of mathematical vocabulary (i.e., Numeration, Geometry-Spatial) will mediate the relation between general vocabulary and specific mathematical skills. This prediction is extended from the '*medium function hypothesis*' which hypothesize that different subcategories of mathematical vocabulary contribute to different mathematical skills uniquely. Specifically, I hypothesize that the Numeration category will partially explain the relation between general vocabulary and arithmetic fluency because numeration items were considered as foundational items (Douglas, 2020). Consistent with what Peng and Lin (2019) found, I also hypothesize that the Geometry-Spatial category will partially explain the relations between general vocabulary and applied mathematical skills (word problem solving and pre-algebra; see Figure 3).

Figure 3. *Extension of Medium Function Hypothesis*



The current study will further contribute to evaluation of the *Function Hypothesis of Language* for mathematics proposed by Peng et al. (2019) by testing the role of mathematical vocabulary and cognitive ability in the relation between general language and mathematical skills and testing effects of different domains of mathematical vocabulary in the relations among general vocabulary, cognitive skills, and different mathematical skills.

Method

Participants

Ethical approval for the research was obtained from the Research Ethics Boards of Carleton University and the University of Winnipeg. Approval was obtained from the Ottawa Carleton District School Board, Manitoba Riel School Division School Board, and Manitoba Seine River School Board. Letters of invitation to participate were sent home with students in three different schools in Ottawa and seven schools in Manitoba. Three of the schools in Manitoba were French Immersion schools. A total of 234 grade 3 students (45.3% male) participated, 87 from Ottawa (49.3% male) and 147 from Manitoba (42.9% male).

Ottawa Sample

Eighty-seven grade 3 students (M_{age} = 8.6 years, *SD* = 3 months, range from 8.1 to 9.2 years) were recruited from three Ottawa elementary schools. All children spoke English fluently, with 58.6% of them having English as their first language. Other first languages included Mandarin (10.3%), Arabic (10.3%), Hindi (2.3%), Russian (2.3%), Turkish (2.3%), and one each of Spanish, Gujarati, Dutch, Japanese, Tigrinya, Hebrew, Urdu, Malayalam, Cantonese, Rankine, Somali, and Cambodian. Of the 81 mothers who provided information about their highest level of education, 2.5% had less than high school, 4.9% had a high-school diploma, 24.7% had a college degree, 39.5% had a university degree, and 28.4% had a post-graduate degree.

Manitoba Sample

Two hundred and nine students (M_{age} = 8.9 years, *SD* = 4.7 months, range from 8.3 to 10.9 years) in grade 3 were recruited from seven Manitoba elementary schools. One hundred and forty-seven of these students had participated in the study in grade 2. Three of the schools were French immersion schools in which all school activities, including announcements and instruction, are in French. Four schools were non-immersion schools where, except for French class, all subjects are taught in English. Among all participants, 92.5% of the students spoke English as their first language, other first languages included German (2%), Chinese-Mandarin (0.7%), Arabic, Yoruba, Russian, Punjabi, Serbian, Korean, and Bosnian. Of the 94% of mothers who provided information about their highest level of education, 2.4% had less than high school, 22.5% had a high school diploma, 25.8% had college degree, 30.6% had a university degree, and 12.4% had a post-graduate degree.

Measures

The list of measures included is shown in Table 3. The complete list of measures used in this study is available on the Open Science Framework at https://osf.io/428hp/.

Table 5. *Measures in the Current Study*

Measures	Ottawa	Manitoba
Cognitive Measures		
Digit Forward[1]	√	√[1]
Visual-spatial Span	√	√
Matrix Reasoning	√	√[1]
Language		
General Vocabulary (PPVT)	√	√
Mathematical Vocabulary (Math Words)	√	√
Mathematical Performance		
Arithmetic Fluency	√	√
Pre-Algebra	√	√
Word Problem Solving	√	√

Notes. Word Problem Solving and Algebra are subtests of the KeyMath III (Connolly, 2007).

[1] Measures collected in Grade 2 for Manitoba sample.

Cognitive Measures

Working Memory. Children's verbal working memory was measured in grade 2 in Manitoba and in grade 3 in Ottawa with two tasks: digit forward span and digit backward span(Ashcraft, 2002). Spatial span, which measures visual-spatial working memory (Alloway et al., 2008), as collected in grade 3 in both Manitoba and Ottawa.

Digit Forward. The task began with a sequence of two digits (e.g., 3-8), to the maximum sequence of 9 digits. In Ottawa in grade 3, the digits sequences were pre-recorded and played to the students, and in Manitoba in grade 2, experimenters read the digits to students. After hearing the stimuli, students were asked to orally recall the digits in sequence, and the experimenters wrote down that sequence in order on the paper (e.g., 3-8). There were two trials for each sequence length. If at least one of the two trials at a given sequence length was recalled correctly, the experimenter increased the sequence length by one digit. Testing was discontinued if the student was unable to correctly recall both trials for a given sequence length. The score was the total number of correct trials that participants recalled.

Spatial Span. The spatial span task was used to assess students' working memory and spatial attention. On each trial, a group of green circles light up one-by-one in a random pattern on the screen of an iPad. Students were instructed to watch the pattern and then tap on the circles in the same order as they lit up. The span-length of the pattern began with two circles and each span was comprised of three trials. After the experimenter demonstrated a practice trial, students were given three more trials of sequences of two locations without any feedback. If a student correctly replicated at least one of the trials for a given span-length, then the next span length increased by one. The task was discontinued if a student were incorrect on all three trials at a given span.

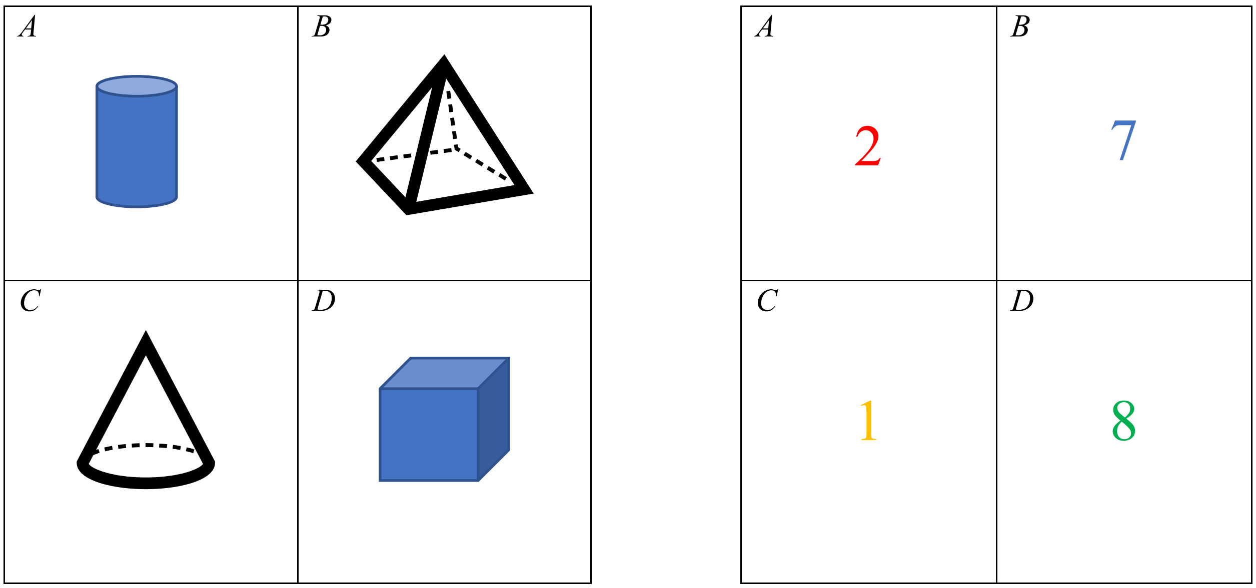
Matrix Reasoning. The Matrix Reasoning subtest of the Weschler Intelligence Scale for Children-Fifth Edition (WISC-5) (Wechsler, 2014) was used to measure children's nonverbal reasoning skills (Ottawa grade 3, Manitoba grade 2). In this task, children were presented with an incomplete grid and asked to select the item that properly completes the matrix, and they were asked either to accurately identify the missing part or detect the incorrect piece of those concepts. There were two practice trials followed by 32 test trials. The experimenter recorded children's responses (correct or incorrect) on a separate answer sheet. Testing was discontinued after three consecutive errors. Score was the number of correct trials.

Language Measures

Receptive Vocabulary. A subset of items from the Peabody Picture Vocabulary Test, Third Edition (PPVT-III) was used to measure receptive English vocabulary (Dunn & Dunn, 1997). During this test, the experimenter said a word and showed the students four pictures. The students' task was to identify the picture that best matched the given word. The vocabulary test is divided into sets of 12 words: five sets (i.e., set 7 to 12) were used. Testing was stopped when children got eight or more words incorrect in a set. Scoring is the number of correct trials. The maximum number of trials was 60.

Mathematics Vocabulary (Math Words). The mathematical vocabulary test was described in detail in the introduction. The Math Words measure contains 40 items in total, with 23 items in Numeration category, 10 items in Geometry and Spatial Sense category, 5 items in Measurement, and 2 items in Data category. The question format was the same format as the receptive vocabulary test: Students were instructed to choose a picture from four options for each mathematical word or expression. The researcher read a math word aloud while the student sees an iPad screen with four images labelled a, b, c, or d (one correct answer, three foils). The student can either point or say the letter of their choice (see examples in Figure 4). Scores are the total number of questions answered correctly out of 40.

Figure 4. *Sample Items from Math Words: Cylinder (left) and Tens column for 7128 (right)*



Note: Actual test questions are not shown here.

Mathematics Measures

Arithmetic Fluency. Arithmetic fluency was a paper-and-pencil measure of children's ability to perform basic math problems (i.e., addition and subtraction) accurately and quickly (Chan & Wong, 2019). Children completed an addition page followed by a subtraction page and multiplication page. Each page was comprised of 60 single-digit problems arranged in three columns for both addition and subtraction (e.g., 4 + 9; 13 - 9). The multiplication questions consist of a total of 60 questions from the 2- ,3- ,4-, and 5-times tables up to 10 and the 6-, 7-, 8-, and 9-times tables to 5 (e.g., 3 x 6; 8 x 1). Students were given one minute per page and asked to write down as many answers as possible without skipping any questions. Scoring is based on the total number of questions answered correctly for each page.

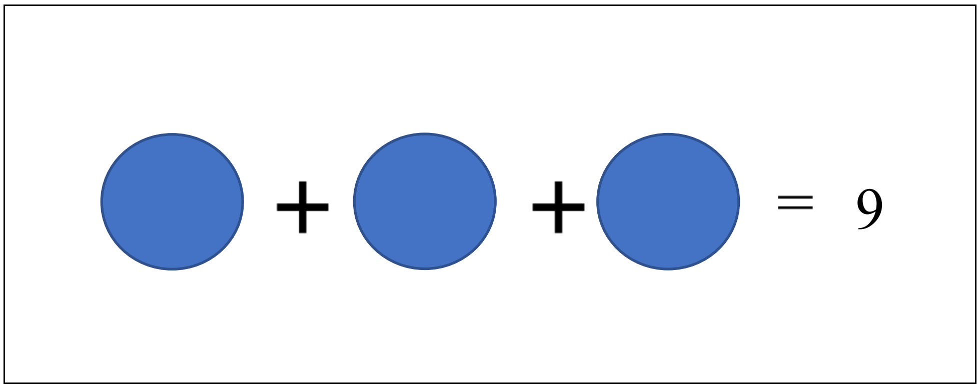
Word Problem Solving. The math word problem solving task was an adapted form of the Applied Problem-Solving subtest of the KeyMath 3rd Edition (Connolly, 2007). Students were shown a picture and asked to respond orally to the corresponding test question (see example in Figure 5). Students were asked to solve a series (i.e., 12 questions) of progressively more difficult mathematics word problems. If a student made three consecutive incorrect responses, the task was discontinued. The scoring was the number of correct responses, with possible scores ranging from 0 to 12.

Figure 5. *Sample Image for Question "There are one hundred thirty-one steps. John is on step sixty. How many more must he climb to reach the top?*

Note: Due to copyright, actual test questions cannot be shown.

Pre-Algebra. The pre-algebra task was adapted from the KeyMath 3rd Edition (Connolly, 2007). Students were shown a picture and asked to respond orally to the corresponding test question (see example in Figure 6). A series of progressively more difficult problems were selected (i.e., 14 grade-appropriate questions). Testing was discontinued after 3 consecutive errors or if they reached the last item. Children were given about 1 minute to respond to each question. Scoring is the total number of correct responses.

Figure 6. *Sample Image for Question "Each circle stands for the same number, three circles equal nine. What number does one circle stand for?*



Note: Due to copyright, actual test questions cannot be shown.

Analyses Plan

Preliminary analyses suggested that there were differences in mathematical language performance among students with different educational experiences. Thus, data were separated by school language status and school location (i.e., Immersion Manitoba, English Manitoba, and English Ottawa). Recall that Immersion students learn mathematics in French, whereas the other two groups learn mathematics in English. The two Manitoba groups share a provincial curriculum. Comparative analyses of all measures with these three groups were conducted first.

Next, the *Function Hypothesis of Language* for Mathematics was tested using mediation analyses after controlling for immersion status and locations. I included four types of measures: cognitive measures (i.e., working memory and matrix reasoning), general vocabulary (receptive vocabulary), mathematical vocabulary, and mathematical performance (i.e., arithmetic fluency, word problem solving, and pre-algebra). Other measures were not used in the current study because they were not directly related to the hypotheses of the study. In terms of mathematical performance, the three outcomes were chosen because of different levels of language processing involved: word problem solving has a high requirement for language processing, whereas arithmetic fluency has a lower requirement for language processing. Pre-algebra has slightly higher language requirement than arithmetic fluency, but it has higher requirement in domain general cognitive skills.

Results

Descriptive Statistics

To examine the performance across groups, distribution of scores within group were explored and then means were compared across three groups (i.e., Immersion Manitoba, English Manitoba, and English Ottawa). The descriptive information for all measures, separated by school immersion status are shown in Table 4. [Note that analyses comparing participants in Manitoba vs. Ottawa are shown in Appendix A]. Based on the skewness values, the measures were normally distributed except for general vocabulary (English Ottawa and Immersion Manitoba), arithmetic fluency (English Manitoba and English Ottawa). General vocabulary performance was negatively skewed in English Ottawa and Immersion Manitoba, but arithmetic fluency was positively skewed in both the English Manitoba and English Ottawa groups. (Outliers were defined as scores with z-scores > |3.29|). Two high outliers were found for arithmetic fluency, one (z= 3.69) in English Ottawa and one (z =3.34) in Immersion Manitoba. Correlational analyses with and without these outliers showed similar patterns of results, and thus none of the data were discarded at this stage.

To estimate internal reliability, Cronbach's alpha was calculated for each measure separately for the three groups. As shown in Table 4, most measures had moderate to excellent reliability (i.e., alpha > 0.70). However, the reliability of Math Words for the Immersion Manitoba and English Manitoba groups were lower, at .63 and .60 respectively. Further analysis of this measure was described in a separate section.

Table 6. *Descriptive Statistics for all Variables*

	N	*Min*	*Max*	*Mean*	*SD*	*Skew*	*Reliability*
Ottawa							
Matrix Reasoning [a]	87	4	22	14.15	4.11	-.50	.92
Digit Forward Span [a]	87	2	14	7.32	1.93	.59	.82
Spatial Span [a]	87	4	15	8.17	2.17	-.21	.95
PPVT [b]	87	2	69	43.99	14.38	-1.08	.98
Math Words[a]	87	9	29	19.26	5.06	-.03	.70
Problem-Solving [a]	87	1	12	6.89	2.67	-.11	.82
Pre-Algebra [a]	87	0	15	8.85	3.60	-.48	.89
Arithmetic Fluency [b]	87	2	130	38.30	24.85	1.37	.92
Immersion Manitoba							
Matrix Reasoning [a]	88	4	23	15.81	3.58	-.49	.79
Digit Forward Span [a]	88	4	13	8.09	2.02	.29	.72
Spatial Span [a]	82	0	18	10.76	3.00	-.51	.79
PPVT [b]	88	3	64	46.42	13.43	-1.29	.98
Math Words [a]	88	6	32	13.88	4.32	1.12	.63
Problem-Solving [a]	88	0	11	5.92	2.46	.12	.83
Pre-Algebra [a]	87	0	14	6.91	3.41	-.05	.91
Arithmetic Fluency [b]	88	3	78	30.76	14.32	.70	.80
English Manitoba							
Matrix Reasoning [a]	59	3	24	16.37	4.24	-.37	.84
Digit Forward Span [a]	59	4	13	8.02	2.29	.33	.76
Spatial Span [a]	52	4	16	10.00	2.84	-.02	.75
PPVT [b]	59	15	64	43.15	12.44	-.39	.97
Math Words [a]	59	6	24	15.93	4.54	-.34	.60
Problem-Solving [a]	59	1	10	5.44	2.14	.23	.81
Pre-Algebra [a]	59	0	14	6.32	3.58	-.21	.92
Arithmetic Fluency [b]	58	4	88	32.28	17.47	1.23	.81

Notes:

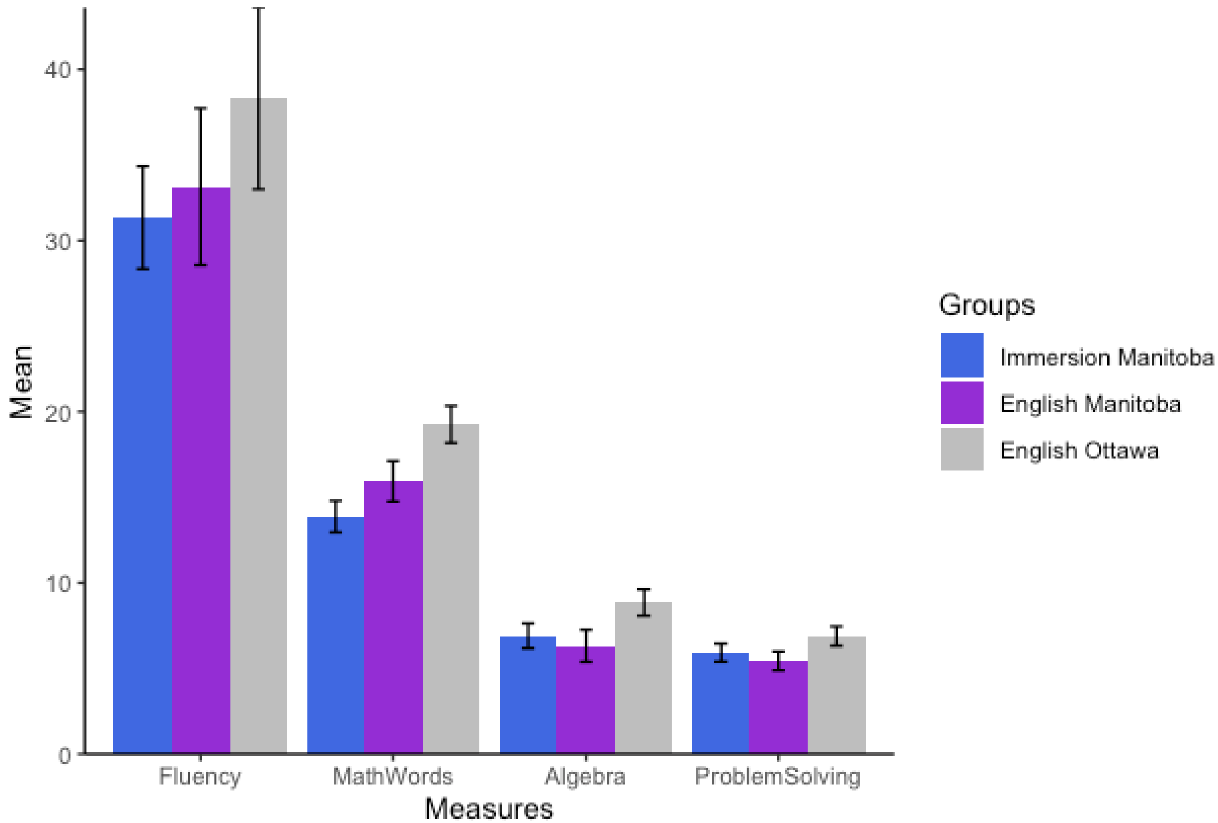
a. Total correct; reliability was calculated based on individual item scores.

b. Total correct for all subsets; reliability was calculated among the subset scores.

Group Comparisons

Matrix reasoning and digit forward span were measured at two different times at the two sites (i.e., grade 2 in Manitoba and grade 3 in Ottawa), and thus scores were converted to standardized (z) scores within groups. Accordingly, comparisons were not made across groups for these measures. For the other measures, however, performance on each measure was analyzed in separate 3 (group: English Ottawa, Immersion Manitoba, English Manitoba) between-groups ANOVAs. General vocabulary, spatial span, and arithmetic fluency were not significantly different across the three groups. Significant differences were found for pre-algebra, $F(2,230) = 10.92$, $p < .001$, problem solving, $F(2,231) = 6.71$, $p = .001$, and Math Words, $F(2,231) = 29.59$, $p < .001$. Means for the math measures across three groups are shown in Figure 7. Tukey's HSD post-hoc tests showed that students in Ottawa performed significantly better than students in Manitoba on math vocabulary, algebra, and problem-solving. There was no difference in pre-algebra and problem-solving between Immersion Manitoba and English Manitoba, but students in the English Manitoba group performed better in math vocabulary than students in the Immersion Manitoba group (See Appendix B). Recall that students in Immersion Manitoba group were taught in French and so they may have been less likely to encounter English math vocabulary words in school.

Figure 7. *Students' Performance on the Mathematics Measures (Whiskers are the 95% Confidence Interval of the Mean).*



In sum, general vocabulary and spatial span were not statistically different across three groups. Performance differences were found in all math outcome measures except arithmetic fluency. In general, the English Ottawa group had significantly higher performance than both Immersion Manitoba and English Manitoba groups. However, there was no significant performance difference found between Immersion Manitoba and English Manitoba groups, except for the Math Words measure, where the English Manitoba group had better performance than Immersion Manitoba group. To control for these mean differences, two contrasts will be included in the correlations and regression analyses: Location (Ottawa vs. Manitoba) and Immersion status (Manitoba English vs. Manitoba Immersion).

Math Words

Because the Math Words measure is central to the current thesis, I examined performance on this task in detail. As shown in Figure 8, the range and pattern of performance was similar in each group even though the means were significantly different (as described in the last section and shown in Figure 7). These mean differences reflected variable performance on some items, however, as shown in Figure 9.

Figure 8. *Comparison of Math Words Measure among Three Groups*

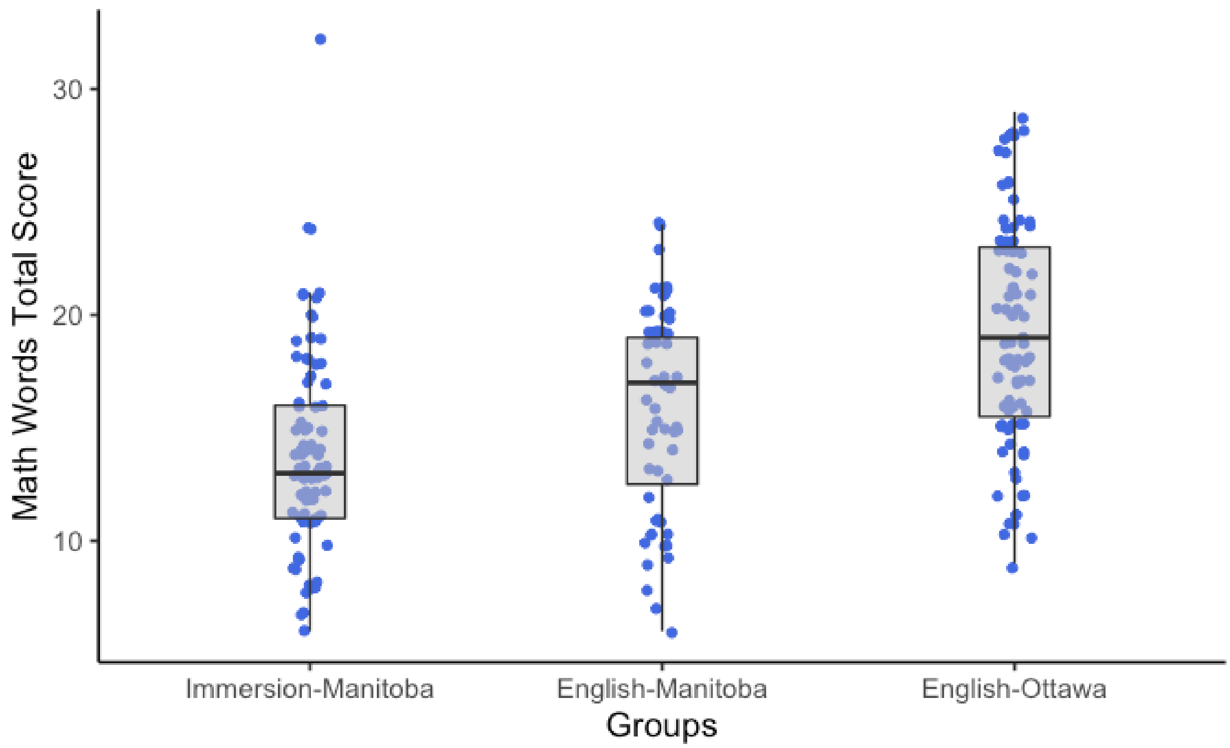
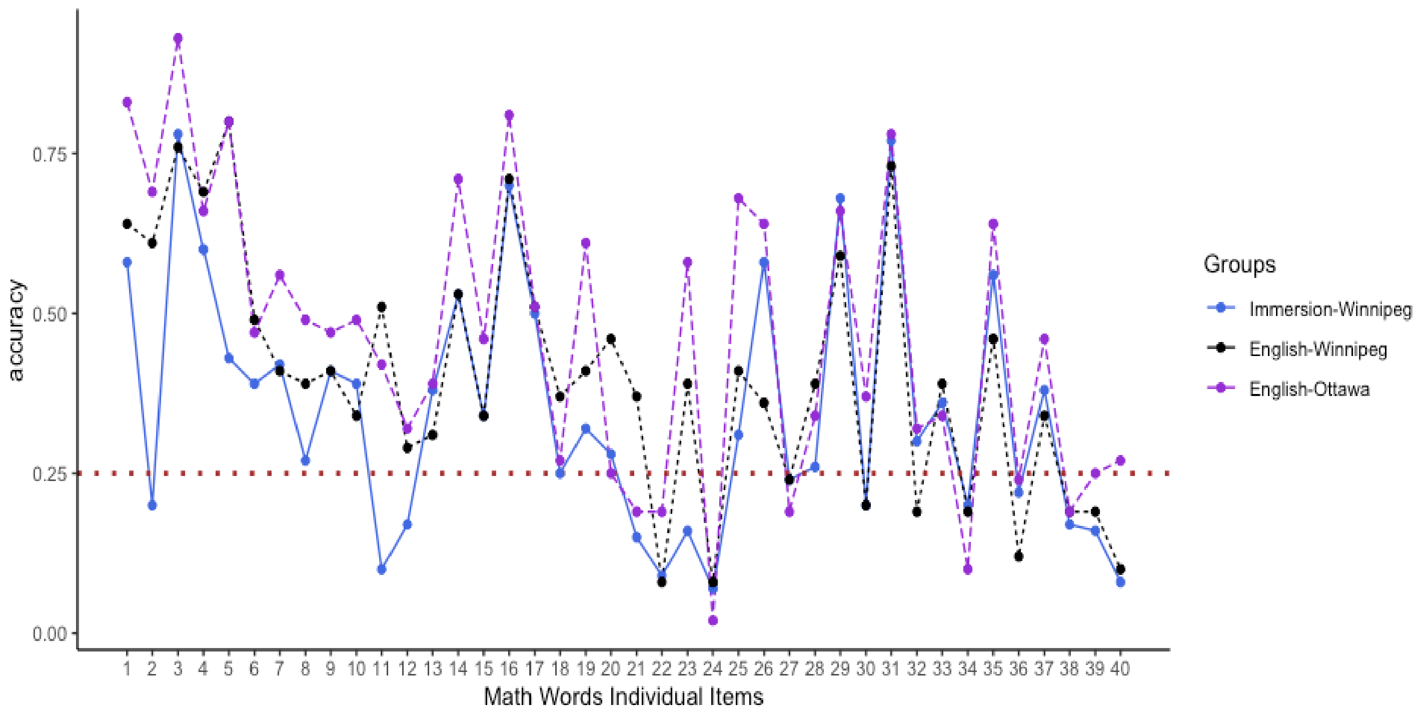


Figure 9. *Accuracy rate of Math Words individual items among three groups.*



Recall that the Math Words measure was multiple choice, with four possible answers for each item. Thus, an accuracy rate lower than 25% for an item may indicate that students guessed. Alternatively, a low accuracy rate may mean that students routinely chose one of the foils on that item. As shown in Figure 9, items 22, 24, 27, 34, 36, 38 and 39 have a low accuracy rate (i.e., rate ≤ 25%) for all three groups of students. All these items were selected from curricula for higher grade levels (i.e., from grade 4 to grade 8). Additionally, items 2, 11, 23 had lower than 25% accuracy rate in the Immersion Manitoba group and higher than 25% accuracy rate in the other two groups. To determine whether differences on specific items were significant across groups, performance on each of the 40 items was analyzed in 3(group) ANOVAs. The means and the results of post-hoc tests are shown in Table 7. For 17 items, the ANOVA results was significant and 15 of these items had significant differences between groups according to Tukey's HSD post-hoc tests. To facilitate comparisons, items were grouped into three blocks in

Table 7 according to the differences in accuracy rate. The first block contains items with low accuracy (25% - 30%) in at least one group and where the performance was not significantly different across the three groups. These items varied in grade levels from 4 to 8, with a median of 5, and thus may have been unfamiliar to most students. The second block of items in Table 7 contains items that are significantly different across three groups. These items varied in grade levels from 1 to 7, with a median of 3. The third block contains items with higher than 30% accuracy rate for all groups and the performance was not significantly different across three groups. These items varied in grade levels from 2 to 7, with a median of 3. (See Appendix C for items accuracy analysis based on grade levels).

Table 7. *Math Words Items by Group*

ID	Item	Grade	*F*	Percentage Correct		
				Immersion Manitoba	English Manitoba	English Ottawa
38	Radius	8	2.84	17.0	18.6	18.6
24	Hundredths column	4	1.49	6.8	8.5	1.7
22	Unit fraction	4	2.32	9.1	8.5	18.6
39	Square root	7	2.77	15.9	18.6	25.4
34	Ratio	6	1.82	20.5	18.6	32.2
36	Prime number	6	1.55	21.6	11.9	23.7
27	Equivalent fraction	5	.29	23.9	23.7	18.6
32	Obtuse angle	5	2.81	29.5	18.6	32.2
18	Factor	4	1.39	25.0	37.3	27.1
28	Multiple	5	1.36	26.1	39.0	33.9
20	Denominator	4	2.51	28.4	45.8	32.2
23	Decimal	4	**25.94*** **	15.9^{a}	39.0^{b}	57.6^{c}
11	Array	3	**18.10*** **	10.2^{a}	50.8^{b}	42.4^{b}
12	Horizontal	3	**3.21***	17.0^{a}	28.8^{b}	32.2^{b}
2	Tens Column	2	**25.94*** **	20.5^{a}	61.0^{b}	69.5^{b}
8	Sum	3	**3.39***	27.3^{a}	39.0^{b}	49.2^{b}
5	Hundreds Column	2	**18.37*** **	43.2^{a}	79.7^{b}	79.7^{b}
40	Integer	7	**6.87*** **	8.0^{a}	10.2^{a}	27.1^{b}
30	Acute angle	5	**6.84*** **	20.5^{a}	20.3^{a}	37.3^{b}
25	Parallel	5	**16.27*** **	30.7^{a}	40.7^{a}	67.8^{b}
19	Key	4	**8.96*** **	31.8^{a}	40.7^{a}	61.0^{b}

14	Perimeter	3	**3.85***	53.4[a]	52.5[a]	71.2[b]
1	Dime	1	**6.19****	58.0[a]	64.4[a]	83.1[b]
3	Cylinder	1	**6.91*****	78.4[a]	76.3[a]	93.2[b]
21	Numerator	4	**5.05****	14.8[a]	37.3[b]	18.6[a]
26	Quotient	6	**4.97****	58.0[a]	35.6[b]	64.4[a]
15	Axes	3	2.74	34.1	33.9	45.8
33	Equilateral triangle	6	.14	36.4	39.0	33.9
13	Product	3	.73	37.5	30.5	39.0
37	Perpendicular	7	1.79	37.5	33.9	45.8
6	Less than	2	.99	38.6	49.2	47.5
10	Difference	3	1.39	38.6	33.9	49.2
9	Ascending	3	.09	40.9	40.7	47.5
7	Mass	3	**3.27***	42.0	40.7	55.9
17	Column	4	.39	50.0	50.8	50.8
35	Negative number	7	2.51	55.7	45.8	64.4
4	One-quarter	2	.75	60.2	69.5	66.1
29	Variable	5	.67	68.2	59.3	66.1
16	Triangular Prism	3	**3.12***	70.5	71.2	81.4
31	Clockwise	3	.29	77.3	72.9	78.0

Note. *** Significant at $p \leq .001$; ** Significant at $p \leq .01$; *Significant at $p \leq .05$.
Means that share superscripts were not significantly different, Tukey HSD, $p < .05$.

For six items in the second block (i.e., Array, Decimal, Tens column, Horizontal, Sum, and Hundreds), students in the Immersion Manitoba group performed significantly worse than the two English groups. For all but one of these items (i.e., Decimal), the other two groups did not differ. One possible explanation for the poor performance in the Immersion Manitoba group was because students had French instruction at school, and they were not familiar with these six English words which were not commonly used in daily life.

For seven items (i.e., acute angle, parallel, key, integer, perimeter, dime, and cylinder), the two Manitoba groups performed significantly worse than the Ottawa English group. The other two items with significant *F* values showed the same pattern (i.e., mass and triangular) but the group differences were not significant in the post-hoc comparisons. The difference found between the two sites were most likely due the different provincial mathematics curriculum requirements.

Finally, for one item (i.e., Numerator), the Manitoba English group performed better than the other two groups, whereas for one item (i.e., quotient), the Manitoba English group performed worse than the other two groups. Overall, the results of the items analysis showed some consistent differences between groups, but a lot of variability among the items.

To determine whether the patterns of relations between Math Words and the other measures would differ if the items which varied across groups were excluded, a new subset score of Math Words was calculated after excluding the 15 items that were significantly different across groups. Correlational analysis was conducted using the subset score for Math Words. All significant correlations remained the same except that the correlation between the subset score for Math Words and the Immersion contrast was not significant, as expected, because the items that differed across these groups were excluded[1]. Because the patterns of correlation did not change when the subset score was used, the complete Math Words score was used in further analyses.

[1] The reliability analysis was conducted with the subset score of Math Words. Cronbach's alpha decreased for all three groups (.37 for English Manitoba, .39 for Immersion Manitoba, and .56 for English Ottawa).

In sum, there was substantial variability in students' performance of Math Words measure among three groups. In general, the students in the Immersion Manitoba had the lowest accuracy rate, and those in the English Ottawa group had the highest accuracy rate. Location and Immersion group status were controlled in the mediation analyses to account for these differences.

Categories of Math Words Measures

Math Words measures included items from four curriculum strands (See Table 1 to Table 4 in the introduction): Numeration (n = 23), Geometry and Spatial Sense (n = 10), Measurement (n = 5), and Data (n = 2). The Measurement and Data categories were not included in the analyses because there were few items from these two categories. Reliability analyses were conducted for Numeration and Geometry-Spatial categories (Table 8). Low Cronbach's alpha may be due to the low number of items in each category and the wide grade range of curriculum grade levels included in each subset.

Table 8. *Cronbach's Alpha for Subsets of Numeration Words and Geometry-Spatial Words*

	Reliability		
	Immersion Manitoba	English Manitoba	English Ottawa
Numeration	.56	.43	.47
Geometry-Spatial	.44	.22	.45

Note. There were 23 items in the Numeration category and 10 in the Geometry-Spatial category.

Correlations

Correlations among the measures for all students are shown in Table 9. Significant correlations were found among all measures except for two demographic measures (i.e., gender and mother's education levels) and the two contrast codes (i.e., location and immersion). Gender was significantly correlated with pre-algebra and arithmetic fluency (boys scored higher than girls). Mother's education level was significantly correlated with matrix reasoning, math vocabulary, pre-algebra, problem-solving, arithmetic fluency, Numeration category, and Geometry-Spatial category. The location contrast was significantly correlated with gender, mother's education level, matrix reasoning ability, Math Words, pre-algebra, problem-solving, arithmetic fluency, and the subset Numeration and Geometry-Spatial categories. The immersion contrast was only correlated with Math Words, pre-algebra, problem-solving, arithmetic fluency, and Numeration subset category. In summary, despite being in the same grade in school, the students in Ontario had higher scores on these measures than the students in Manitoba. Gender, mother's education level, and the location and immersion status contrasts were controlled in the mediation analyses.

As shown in Table 9, correlations between Math Words and math outcome measures (i.e., pre-algebra, problem-solving, and arithmetic fluency) were all above .44, and all math outcome measures were more highly correlated (above .53) with each other than with other measures. In addition, general vocabulary was not highly correlated (r =.38) with Math Words, but both general vocabulary and Math Words were highly correlated with math outcome measures. The two Math Words subcategories – Numeration and Geometry-Spatial -- were moderately correlated with all math outcome measures (i.e., higher than .38), except that the correlation between Geometry-Spatial category and arithmetic fluency is somewhat lower at .30. Furthermore, the Numeration and Geometry-Spatial subcategory scores were significantly correlated (.46), which indicated that the two subcategories of Math Words shared some variance of, but they also captured some unique variance in each subcategory.

Table 9. *Correlations Among Measures for All Students (N=234).*

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Gender	-												
2. Mother's education	-.13	-											
3.Location Contrast[a]	-.06	.26**	-										
4.Immersion Contrast[b]	-.05	-.17**	.12										
5. Digit forward (Z)	-.02	-.02	.00	-.01	-								
6. Matrix reasoning (Z)	.07	.16*	.00	.06	.19**	-							
7. Spatial span	-.07	.06	.04	-.11	.18**	.31**	-						
8. General Vocabulary	.08	.10	-.04	-.10	.32**	.15*	.24**	-					
9. Math Words	-.10	.30**	.43**	.20**	.27**	.30**	.23**	.38**	-				
10. Pre-algebra	-.13*	.31**	.29**	-.03	.32**	.42**	.39**	.41**	.54**	-			
11. Problem-solving	-.10	.26**	.22**	-.05	.37**	.34**	.39**	.40**	.53**	.69**	-		
12. Arithmetic fluency	-.21**	.30**	.16**	.06	.17**	.27**	.28**	.14*	.44**	.53**	.51**	-	
13. Geometry MW[c]	-.07	.27**	.38**	.06	.16*	.17**	.17**	.33**	.73**	.41**	.38**	.29**	-
14. Numeration MW[d]	-.10	.22**	.31**	.25**	.26**	.28**	.23**	.29**	.88**	.51**	.47**	.43**	.46**

Note. $p < .05$*; $p < .01$**.

[a]. Contrast coding of comparison between English Ottawa and two Manitoba groups (Immersion and English).

[b]. Contrast coding of comparison between Immersion Manitoba and English Manitoba

Correlation Comparisons

Three groups of correlation comparisons were conducted (Table 10). First, comparisons were conducted for correlations between general vocabulary and Math Words with the same math outcome measure, for example, I compared the correlations between general vocabulary and Math Words with pre-algebra. As shown in the first row of Table 10, the correlation between Math Words and each outcome was higher than the correlation between general vocabulary and that outcome. This pattern indicated that the Math Words measure captured more variance in math outcomes than general vocabulary did.

Table 10. *Comparisons between Correlations for Vocabulary Measures with Math Outcomes using Hotelling's t.*

	Math outcomes *(df* = 231)					
	Pre-algebra		Problem-solving		Fluency	
Comparisons	*t*	*p*	*t*	*p*	*t*	*p*
General vocabulary vs. Math Words	-2.19*	.030	-2.16*	.032	-4.56**	<.001
Math Words vs. Numeration	1.11	.268	2.20*	.029	.35	.729
Math Words vs. Geometry-spatial	3.20***	.002	3.66***	<.001	3.39***	<.001
Numeration vs. Geometry-spatial	1.75	.082	1.53	.129	2.28*	.023

Note. * $p < .05$; ** $p < .01$; *** $p < .001$

Second, correlation comparisons were conducted for correlations between each of the math outcome variables and the total Math Words measure versus the subcategories (Numeration, Geometry-Spatial) for each of the math outcome variables. As shown in the second row of Table 10, the correlation was higher between Math Words and problem solving than between Numeration Words and problem solving, but there was no difference in these correlations for pre-algebra or arithmetic fluency. In contrast, the correlation was higher between Math Words and all three outcomes than between the Geometry-Spatial Words and all three outcomes. These findings indicate that the Math Words measure shared more variance with all math outcomes than did Geometry-Spatial subcategory, and except for the problem-solving, Math Words and Numeration subcategory may share the same variance with pre-algebra and arithmetic fluency.

Third, correlation comparisons were conducted for correlations between Numeration and Geometry-Spatial Words with the same math outcome measures. As shown in the fourth row of Table 10, only one significant difference was found. The correlation between Numeration and arithmetic fluency was higher than that between Geometry-Spatial subcategory and arithmetic. In general, therefore, the relations between subcategories of Math Words and math outcomes did not differ, except for arithmetic fluency.

In summary, the comparisons indicate that the complete Math Words measure had a stronger correlation with math outcomes than general vocabulary, and that the Math Words measure shared more variance with math outcomes than did the subcategories. Moreover, the correlation comparison between Numeration Words and Geometry-Spatial Words indicated that these two subcategories tap into different knowledge, which means different subcategories may have different influence on students' math performances.

Having established relations among mathematical vocabulary, cognitive measures, and math outcome measures, I tested the three hypotheses with mediation analyses while controlling for students' gender, mothers' education level, and the two control contrasts for location and immersion status.

Mediation Analyses

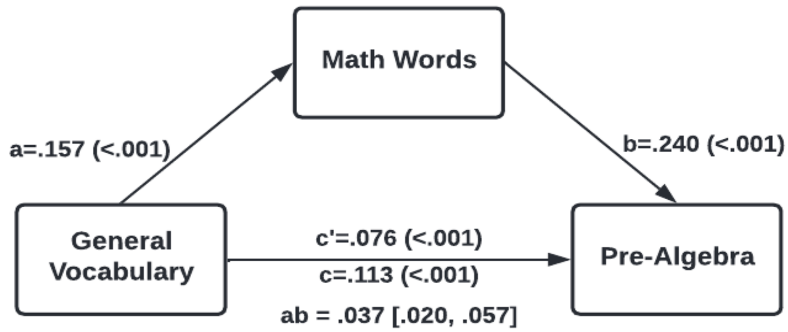
For mediation to occur, the potential mediator, predictor, and the outcome must be significantly correlated with each other. If the indirect effect of the predictor on the outcome through mediator is significant, and the direct effect of the predictor is reduced, then mediation has occurred. Using the PROCESS tool in SPSS, the significance of the indirect effect was assessed using confidence intervals estimated with bootstrapping methods.

Hypothesis 1: For students in grade 3, mathematical vocabulary mediates the relation between general vocabulary and mathematical skills, after controlling for gender, mothers' education, the location contrast, and the immersion contrast.

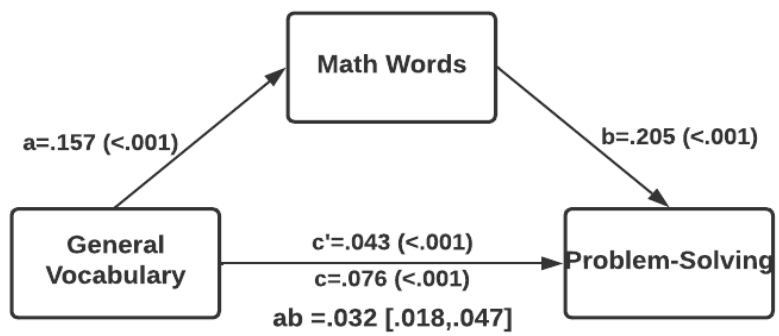
As shown in Figure 10, mathematical vocabulary mediated between general vocabulary and all three outcomes. The confidence intervals did not include zero and significance levels using the Sobel test were all less than .05 (See Table 11 and Figure 10). Mathematical vocabulary partially mediated the relation between general vocabulary and two applied mathematical skills (pre-algebra and problem-solving). In addition, the relation between general vocabulary and arithmetic fluency is completely mediated by mathematical vocabulary. The coefficients for the indirect effect of vocabulary on the outcomes are shown in Table 11.

Figure 10. *Tests of Mediation of General Vocabulary with Mathematical Vocabulary.*

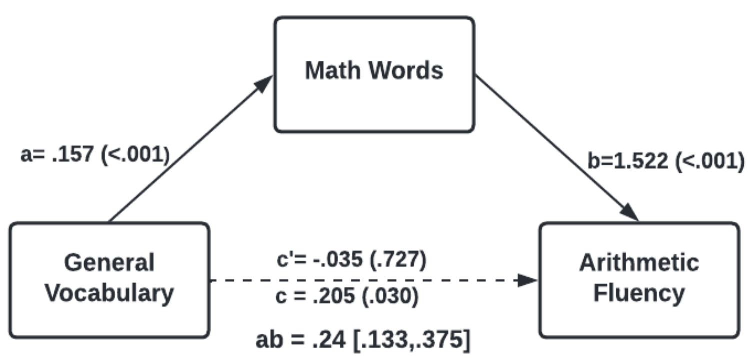
a)



b)



c)



Note. Dotted line represents non-significant direct effect; ab is the coefficient of the indirect effect of general vocabulary on math outcomes. The 95% confidence interval for the indirect effect is a BCa bootstrapped CI based on 5000 samples. Covariates were controlled in all models, gender (only significant in model c, p =.01), mothers' education (significant in model b, p =.012; and model c, p <.001), location contrast (significant in model a, p =.023; and model b, p= .002), and immersion contrast (not significant in any models).

Table 11. *Summary of Indirect Effects of all Mediation Models*

		Mediation models					
Predictors	Mediators	Pre-algebra		Problem-solving		Arithmetic fluency	
		B	95% CI	*B*	95% CI	*B*	95% CI
General vocab	Math Words	**.037**	[.021, .056]	**.032**	[.019, .047]	**.234**	[.134, .372]
Math Words	Digit forward	**.031**	[.007, .064]	**.026**	[.005,.053]	.087	[-.029, .243]
	Matrix reasoning	**.037**	[.008, .077]	.015	[-.004, .040]	.062	[-.102, .248]
	Spatial span	**.023**	[.000, .054]	**.021**	[.002, .046]	.121	[-.001, .286]
General vocab	Numeration	**.027**	[.018, .054]	**.019**	[.009, .031]	**.158**	[.068, .278]
	Geometry-Spatial	.008	[-.003, .020]	.007	[-.002, .016]	.036	[-.034, .114]

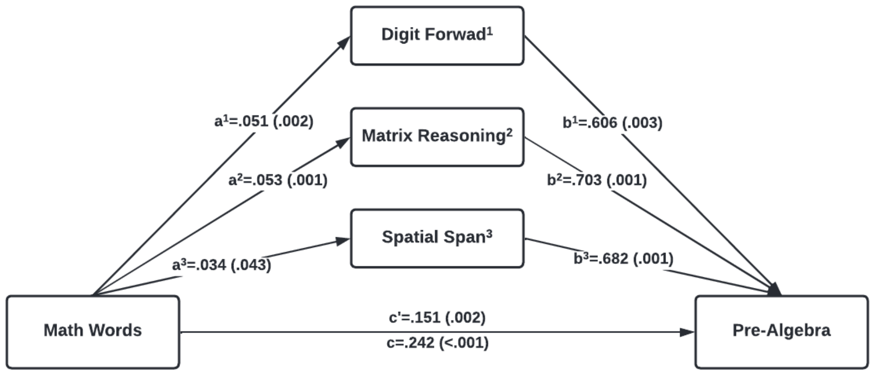
Hypothesis 2: Students' cognitive skills mediate the relation between mathematical vocabulary and math outcomes, after controlling for general vocabulary, gender, mothers' education, location contrast.

As shown in Table 11 and Figure 11a, the effect of mathematical vocabulary on pre-algebra was partially mediated by digit forward, matrix reasoning, and spatial span. In the second model (Figure 11b), the effect of Math Words on problem solving was partially mediated by digit forward and spatial span, but matrix reasoning did not mediate the relation between mathematical vocabulary and problem-solving. In the last model (Figure 11c), the path from mathematical vocabulary to the spatial span, and the path from spatial span to arithmetic fluency were significant, but the indirect effect through spatial span is not significant.

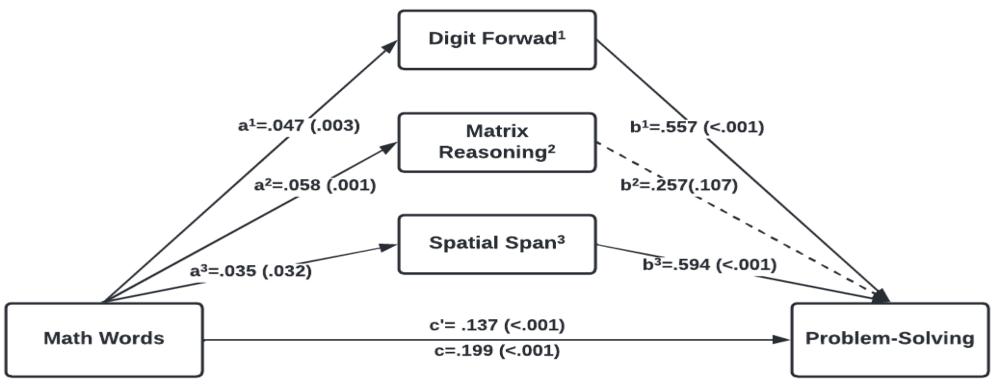
In sum, spatial span and digit forward significantly mediated the relation between mathematical vocabulary and applied mathematical outcomes (pre-algebra and problem-solving), but matrix reasoning only mediated the relation between mathematical vocabulary and pre-algebra. Although spatial span did not mediate the relation between mathematical vocabulary and arithmetic fluency, it was a successful predictor of arithmetic fluency. In addition, Math Words measure was a predictor of all cognitive measures, and it still predicted math outcomes after accounting for the mediators.

Figure 11. *Models of Mathematical Vocabulary as a Predictor of Math Outcomes, as Mediated by Three Cognitive Measures (digit forward, matrix reasoning and spatial span).*

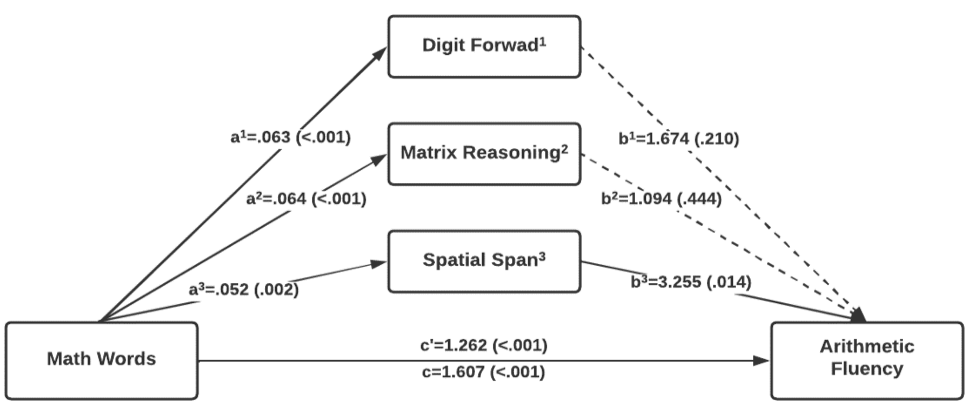
a)



b)



c)



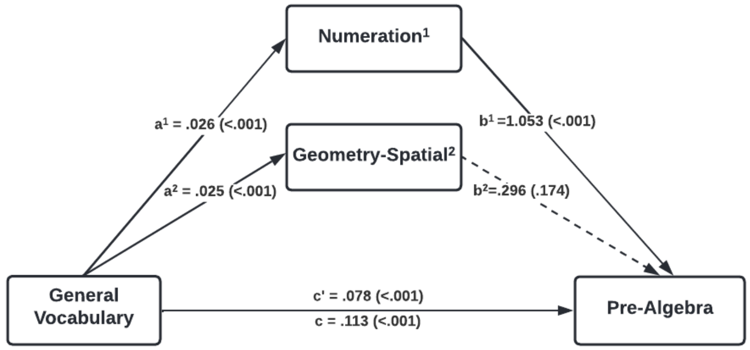
Note. All three cognitive measures were standardized into z score to facilitate the comparison. Dotted lines represent non-significant paths. The 95% confidence interval for the indirect effect is a BCa bootstrapped CI based on 5000 samples. General vocabulary (p <.001 in model a; p =.039 in model b), gender (p = .043 in model 3), mothers' education (p = .021 in model a; p = .027 in model b; p = .006 in model c), and location contrast (p = .004 in model a) were controlled in all mediation models.

Hypothesis 3: The mathematical vocabulary subcategories mediate the relation between general vocabulary and math outcomes, after controlling for gender, mothers' education, location contrast, immersion contrast.

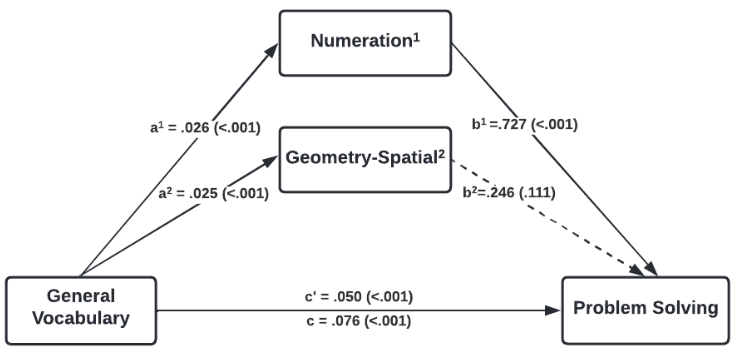
As shown in Table 11 and Figure 12, Numeration Words partially mediated the effect of general vocabulary on pre-algebra, and partially mediated the effect of general vocabulary on problem solving. Numeration Words fully mediated the effect of general vocabulary on arithmetic fluency. The indirect effect of general vocabulary through Numeration Words was significant. In contrast, the Geometry-Spatial category did not mediate the relations between general vocabulary and math outcomes.

Figure 12. *Models of General Vocabulary as Predictor of Math Outcomes, Mediated by Mathematical Vocabulary Subcategories (Numeration and Geometry-Spatial).*

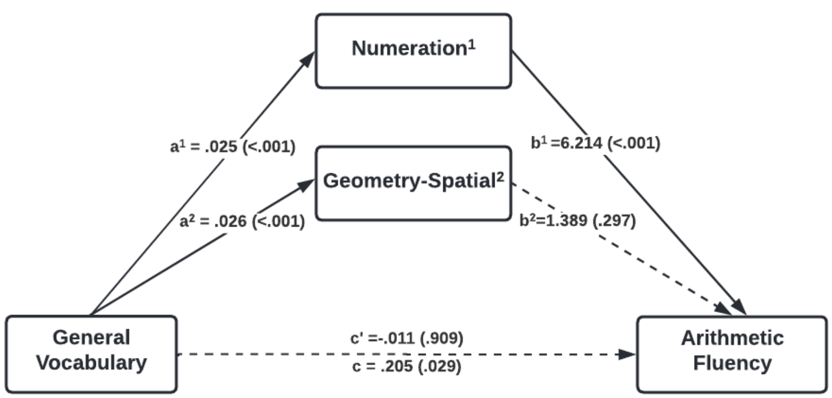
a)



b)



c)



Note. Within each subcategory, z scores were calculated to equate the scales to facilitate the comparison. Dotted lines represent non-significant paths. The 95% confidence interval for the indirect effect is a BCa bootstrapped CI based on 5000 samples. Gender (p=.041 in model 1), mothers' education (p= .004 in model 1; p=.045 in model 3), location contrast (p= .001 in model 2; p<.001 in model 3), and immersion contrast were controlled in all mediation models.

The results of these analyses were very similar to those for Hypothesis 1. The same predictor (general vocabulary) and the mediator (Math Words) were used in both hypotheses, except the subcategories of Math Words were used in Hypothesis 3 instead of the whole set. The explained variance in models testing Hypothesis 3 were close to the R^2s in models of Hypothesis 1 (see Table 12). It is possible that Numeration Words was accounting for the same variance as the whole Math Words measure. I re-ran three models with only Numeration Words as the mediator and compared the R^2 with the original models (see Table 12). The results confirmed that Numeration Words basically carried the same variance as the whole set of Math Words. Indicating that the Numeration items captured the same underlying construct as the whole Math Words measures.

Table 12. *Summary of R2 of all Mediation Models*

Hypothesis	Predictor	Mediators	Mediation Models R^2		
			Pre-algebra	Problem-solving	Arithmetic fluency
1	General vocabulary	Math Words	.394	.361	.243
2	Math Words	Digit forward Matrix reasoning Spatial span	.510	.475	.300
3	General vocabulary	Numeration Geometry-spatial	.413	.349	.246
3[a]	General vocabulary	Numeration	.407	.341	.242

Note. [a] indicate the model that contains only Numeration Words as the mediator in the mediation analyses.

Discussion

Mathematics is an essential foundation for student success in disciplines such as science, technology, and engineering (Baker & Galanti, 2017; English, 2017). Many different skills, such as language, working memory, and nonverbal reasoning skill, are involved in mathematics (Morsanyi et al., 2018; Peng et al., 2020; Raghubar et al., 2010). Therefore, it is important for educators to understand how language and other skills influence mathematical learning. The first goal of the present study was to examine the relations among general vocabulary, mathematical vocabulary skills, and different mathematical skills (i.e., pre-algebra, problem-solving, and arithmetic fluency) for students in grade 3, and second goal was to determine whether domain-general cognitive skills (i.e., vocabulary, digit forward, matrix reasoning, and spatial span) play a role in this relation. The third goal was to test whether subcategories of Math Words (i.e., Numeration and Geometry-Spatial) mediate the relation between general vocabulary and mathematical skills differently. I will discuss the results of the analyses proposed in the introduction and the limitations of this work and propose future research to extend our understanding of mathematical vocabulary.

Correlations

Mathematical vocabulary was significantly correlated with all math outcomes, ranging from .44 with arithmetic fluency to .54 with pre-algebra. In general, this pattern is consistent with the meta-analysis by Lin et al. (2021) in which the average correlation between mathematical vocabulary and mathematical performances was .49. Moreover, the current study found the correlations between subcategories of Math Words -- Numeration and Geometry-Spatial -- and math outcomes were also all significant, ranging from .29 (between Geometry-Spatial Words and arithmetic fluency) to .51 (between Numeration Words and pre-algebra, see

Table 9). In summary, the present research found correlations which were consistent with previous findings showing that mathematical vocabulary is more strongly related to measures of mathematical performances than is general vocabulary (Lin et al., 2021; Powell et al., 2017; Purpura et al. 2011, 2017, Purpura & Logan, 2015).

Mediation Analyses

Hypothesis 1

The *Function Hypothesis of Language for Mathematics* (Peng et al., 2020) suggested that mathematical language serves both a *medium function* and a *thinking function* in students' mathematical learning. Mediation analyses were conducted to test three hypotheses that arose from Peng et al.'s proposed model. Hypothesis 1 examined the *medium function* by testing the mediation effect of mathematical vocabulary for the relation between of general vocabulary and mathematical performance for three outcomes, arithmetic fluency, problem solving, and pre-algebra. After controlling for students' gender, mothers' education level, and both location and immersion contrasts, students' mathematical vocabulary skills partially mediated the relation between general vocabulary and two applied mathematical skills (pre-algebra and problem-solving; see Figure 10). These findings support the *medium function hypothesis* for these outcomes, that is, partial mediation for students at this level of skill (Peng et al., 2020).

In contrast to the results for pre-algebra and problem solving, students' mathematical vocabulary fully mediated the relation between general vocabulary and arithmetic fluency. Why was the mediation complete for arithmetic fluency but only partial for the other two measures? A plausible explanation for the difference lies in the language requirements of the math outcome measures. Students needed knowledge of both general vocabulary and mathematical vocabulary to interpret the questions on the applied mathematical measures (i.e., problem-solving and pre-algebra). In contrast, the arithmetic fluency task was fully symbolic and thus no language processing was required in the task other than the oral instructions at the beginning. Thus, the requirement for interpretation and use of language processes was minimal, although specific mathematical vocabulary may be retrieved for accessing knowledge about the operations, such as add, subtract, and multiply. Moreover, Powell et al. (2017) indicated that mathematical vocabulary is more than just a language measure, it is also a measure of mathematical knowledge. Consistent with this view, mathematical vocabulary was more strongly correlated with mathematical performances than was general vocabulary. In sum, the analyses supported Hypothesis 1 and the medium *function* hypothesis (Lin et al., 2021).

Hypothesis 2

The analyses to test Hypothesis 2 examined the *thinking function* by testing the mediation effect of three cognitive skills (i.e., digit forward, matrix reasoning, and spatial span) in the relation between mathematical vocabulary and math performances. After controlling for students' gender, mothers' education level, general vocabulary, and both location and immersion contrasts (i.e., group mean differences), Hypothesis 2 was supported because the results showed that cognitive skills partially mediated the relations between mathematical vocabulary and applied mathematical skills (i.e., pre-algebra, and problem-solving). Specifically, all three cognitive skills partially mediated the relation between mathematical vocabulary and pre-algebra. Digit forward and spatial span, but not matrix reasoning, mediated the relation between mathematical vocabulary and problem-solving. In contrast to these results, cognitive skills did not mediate the relations between mathematical vocabulary and arithmetic fluency.

The different mediation effects for the three outcome tasks was because different math outcome tasks required different cognitive skills. Both digit forward and spatial span measured students' working memory. Matrix reasoning measured students' abstract nonverbal reasoning skill. Both working memory and nonverbal reasoning skill explained the part of the relation between mathematical vocabulary and pre-algebra. But only working memory explained part of the relation between mathematical vocabulary and problem-solving. One possible explanation for this difference is that the pre-algebra task included nonverbal symbols in the questions which taps into students' nonverbal reasoning skill (e.g., see Figure 6 in the Introduction). But problem-solving did not include any nonverbal symbols, which may not require students' nonverbal reasoning skills. However, working memory was required to maintain the task-relevant information and perform the task for these two applied mathematical skills.

Even though none of the cognitive skills mediated the relation between mathematical vocabulary and arithmetic fluency, mathematical vocabulary and spatial span both predicted unique variance in students' arithmetic fluency (see Figure 11c). The results were consistent with previous findings which indicated that children's working memory (Berg, 2008; DeStefano & LeFevre, 2004) and mathematical vocabulary (Ünal et al., 2021; Xu et al., 2022) are strongly related to arithmetic fluency. But the results showed that cognitive skills did not explain the relation between mathematical vocabulary and arithmetic fluency.

Overall, the results supported the *thinking function* hypothesis by showing that working memory and nonverbal reasoning ability did not explain all variance in the relation between mathematical vocabulary and higher-order math performance (i.e., pre-algebra and problem-solving), and did not mediate the relation between mathematical vocabulary and arithmetic fluency (Peng et al., 2020; Peng & Lin, 2019). Thus, mathematical vocabulary may be actively involved in students' comprehension and thinking processes in these mathematical tasks.

Hypothesis 3

Hypothesis 3 is an extension of the *medium function* hypothesis: On this view, subcategories of mathematical vocabulary are differentially related to mathematical skills (Peng et al., 2020). Mediation effects of Numeration and Geometry-Spatial word knowledge in the relation between general vocabulary and math outcomes were tested. Students' gender, mothers' education level, and both location and immersion contrasts were controlled in the mediation analyses. In general, the results supported Hypothesis 3. Numeration partially mediated the relation between general vocabulary and applied mathematical skills (pre-algebra and problem-solving), also, Numeration fully mediated the relation between general vocabulary and arithmetic fluency. However, Geometry-Spatial did not mediate any of the relations between general vocabulary and mathematical performances.

Possible explanations for differences in the predictive relations of the two Math Words subcategories is that the Numeration category included 23 items which is more than half of the Math Words measure, and 14 out of 23 items were selected from lower grade levels (grade 2 to grade 4). In contrast, the Geometry-Spatial category included only 10 items, and 6 of the items were selected from higher grade levels (grade 5 to grade 8). Thus, Numeration Words may fully mediate the relation between general vocabulary and arithmetic fluency because numeration words (e.g., sum, product, multiple) may be retrieved by students during the calculation process. The partial mediation effect of the Numeration Words for applied mathematical skills was because both pre-algebra and problem-solving required of more than just fundamental numeration knowledge. In summary, the Numeration Words was a better measure which shared the accurate variance with the mathematical skill than the Geometry-Spatial Words for grade 3 students.

The results of the present analyses were different from those of Peng and Lin (2019) who found that the Geometry category explained the relation between general vocabulary and problem-solving, also Numeration category did not mediate the relation between general vocabulary and arithmetic fluency. The difference may be explained by comparing the mathematical vocabulary measures between the current study and Peng and Lin (2019). Peng and Lin tested grade 4 students, and they included words from grade 3 to grade 5 in their mathematical vocabulary measure, but the Math Words measure used in the present research included words from grade 1 to grade 8 to test grade 3 students' knowledge of mathematical vocabulary. Thus, the Math Words measure had a wide range of mathematical vocabulary which may increase the variability in students' performances. Moreover, Eight of the 23 items from the Math Words Numeration category overlapped with those of Peng and Lin (n = 34), and four of 10 items from Geometry-Spatial category overlapped with those of Peng and Lin (n = 30). The words selections were different from the measure used in Peng and Lin (2019), also, Chinese students had better performances than students in the current study. For example, both studies included a 3-minutes timed arithmetic fluency measure, Chinese grade 4 students had average 104 correct answer (total 160 questions) in arithmetic fluency, and students in the current study had average 33 correct answer (total 180 questions). The different word selection and different students' performances may be due to different education requirements between Canada and China. In the current study, the Numeration category may be a more valid measure of mathematical knowledge for grade 3 students comparing to the Geometry-Spatial category.

In sum, the mediation analyses confirmed that mathematical vocabulary mediated the relation between general vocabulary and mathematical performances (Fuchs et al., 2015; Powell et al., 2017; Ünal et al., 2021), which suggested that mathematical vocabulary may contribute to mathematics learning directly and facilitate the learning of advanced mathematical skills. Moreover, the results built on existing evidence of mathematical vocabulary categorization (Lin et al., 2021; Monroe & Panchyshyn, 1995; Peng & Lin, 2019; Powell & Nelson, 2017; Rubenstein & Thompson, 2002) which suggested that mathematical vocabulary may have content-specific subcategories that are closely related to specific mathematical content, but more importantly, the specific mathematical vocabulary matters only if it reflected the specific mathematical knowledge that was required to perform the tasks.

Limitations and Recommendations for Future Research

The current study provided important information about mathematical vocabulary and its relations with mathematical outcomes. Nonetheless, several limitations of the current study should be noted.

First, the reliability of the Math Words measure was low, suggesting that did not reflect a unitary latent construct. Compared with the mathematical vocabulary measures used in other studies, the Cronbach's alpha of mathematical vocabulary measure was .86 in Fuchs et al. (2015) and .85 in Powell and Nelson (2017). The Math Words measure in the present study was originally designed for students in grades 4 and 6, the Cronbach's alpha was .76 (Douglas, 2020), so it is possible that the Math Words measure may not represent grade 3 students' knowledge of mathematical vocabulary accurately. Nevertheless, the correlations between the Math Words measure and math outcome measures were quite similar to the average of .49 reported by Lin et al. (2021). Thus, the Math Words measure captured some variance that was shared with students' mathematical performances, but it contained items that may not accurately reflect the mathematical knowledge in grade 3, and may increase the variability in students' performances.

Second, the Math Words measure had several other limitations: 1) a small number of items included, 2) items were selected from a wide grade range (grade 1 to grade 8), and 3) items were not systematically selected. In consequence, I did not have equal number of items in Numeration and Geometry-Spatial categories, which may cause bias in the results. For example, I only found a mediation effect of Numeration category, but not the Geometry-spatial category, and this result may be due to the unequal numbers of items and unequal representations of grade level in the subcategories. The numeration category included 23 items which is more than half of the Math Words measure, and 14 out of 23 items (61%) were selected from lower grade levels (grade 2 to grade 4). In comparison, the Geometry-Spatial category included only 10 items, and 4 (40%) of the items were selected from lower grade levels. In addition, comparing Hypothesis 1 with Hypothesis 3 with the same control variables (see Table 12), the Numeration category explained the same variance as the complete Math Words measure, which implied that the Numeration category was most relevant to what students were learning in grade 3.

Overall, the specificity of mathematical vocabulary may be supported, but the results should be interpreted with caution. I only found the significant mediation effect of Numeration category, and it may be because the Numeration Words captured most information about students' mathematical vocabulary knowledge in the current study. Those Geometry-Spatial Words (e.g., radius, denominator) may be as important as the Numeration Words (e.g., less, sum), but the mathematical outcome measures in the current study may not require knowledge of Geometry-Spatial Words. Moreover, Math Words (Douglas, 2020) was not designed to test the specificity of the mathematical vocabulary originally, also the Math Words was created without accessing to the other work (Peng & Lin, 2019; Powell & Nelson, 2017). In future designs of mathematical vocabulary measures, researchers could replicate what Powell and Nelson did, that is consider all possible mathematical vocabulary in the mathematical curriculum. Also, to study the specificity of mathematical vocabulary, in the future studies researchers should create balanced domain-specific subcategories of mathematical vocabulary (e.g., radius, circle, length, centimeter), and researchers should use tasks that require the subcategories of mathematical vocabulary knowledge (e.g., geometry, measurement).

Third, students performed differently from each other across individual items on the Math Words measure. In general, students in the Immersion Manitoba group performed worse than those in both the English Manitoba and English Ottawa groups. The difference may be due to the different language instruction students experienced at school and at home. The Math Words measure used English words, which may not capture the full picture of Immersion students' mathematical vocabulary knowledge, thus, the poor performance was unsurprising. Also, I did not have information about how much exposure of English mathematical vocabulary that Immersion students had in school or in their daily life. In other words, the performance of Math Words measure in Immersion Manitoba group may be confounded by the difference in language. Therefore, the immersion contrast was controlled in the study, and the patterns that were identified in this study should not be influenced by the group difference.

A final limitation of the current research was that, despite the original goal of performing an item response theory (IRT) analysis with 2-PL model for the Math Words measure, the IRT was not applicable for the current data. To get an acceptable accuracy, a minimum sample size of 750 was suggested for a 10-item measure in 2-PL model (Sahin & Anil, 2017). Moreover, to perform the IRT, unitary construct of the test was assumed (Hambleton et al., 1991), but the

Math Words measure may not be a unitary construct. Therefore, the IRT was not useable with the current study. IRT could provide information about individual variance, which assumes that person with similar laten traits will obtain similar scores on the different test built to evaluate the same construct. (Hambleton et al., 1991; StataCorp, 2017). In the future studies with larger sample size, researchers could use IRT to better understand the individual variance in performance of each subcategory of Math Words measure.

Conclusion

In the current study, I found that grade 3 students' knowledge of mathematical vocabulary was related to their mathematical performance. These results supported the view that language, specifically vocabulary, plays an important role in mathematics learning (LeFevre et al., 2010; Lin et al., 2021; Peng & Lin, 2019). Teachers communicate ideas through language and students share ideas through language. Students' understanding of content-specific language (i.e., mathematical vocabulary) can facilitate the learning of mathematics. Another contribution of the present study is to confirm that mathematical vocabulary may be categorized into domain-specific subcategories (Fuchs et al., 2015; Peng & Lin, 2019; Powell & Nelson, 2017; Purpura & Reid, 2016). However, specificity is associated with specific mathematical knowledge, and it can play a significant role if the specific mathematical words were relevant to the task. Thus, specificity would be important when it contributes to students' learning about specific mathematical skills, such as geometry, and fraction.

APPENDICES

Appendix A. Comparative Analyses between Two Sites.

An initial analysis was done comparing the data between the two provinces, Ontario and Manitoba. The measures were normally distributed except for general vocabulary (Ottawa), and arithmetic fluency (Manitoba and Ottawa) (see Table 1). Outliers were defined as scores with z-scores > |3.29|). One outlier was found for arithmetic fluency in Ottawa and one for Math Words in Manitoba, two outliers were found for general vocabulary in Manitoba and two for matrix reasoning in Manitoba, and three outliers were found for arithmetic fluency in Manitoba. Comparisons of means of all measurements between two sites were conducted. Results showed that the mean difference between Ottawa and Manitoba was statistically significant for Math Words (M_o 19.26 vs. M_m 14.7), F (1, 232) = 51.02, p < .001; Problem-solving (M_o 6.89 vs. M_w 5.73), F (1, 232) = 12.06, p <.001; Pre-Algebra (M_o 8.85 vs. M_w 6.67), F (1, 231)= 20.86, p <.001); and Arithmetic Fluency ((M_o 38.30 vs. M_w 32.05), F(1, 227) = 5.57, p = .019.

Appendix B. Post-hoc Comparison of Three Math Outcomes among Three Groups.

Tukey's HSD post-hoc comparison was conducted for Math Words, problem-solving, and pre-algebra (See Table 13). The results showed the difference was significant across three groups for Math Words where English Ottawa had better performance than the other two group, and Immersion Manitoba had worse performance than the other two groups. For both pre-algebra and problem-solving, English Ottawa had better performance than the other two groups, but there was no performance difference between English Manitoba and Immersion Manitoba.

Table 13. *Post-hoc Comparison of Three Math Outcomes among Three Groups.*

Measurements	Group comparison		Mean difference	Sig.
Math Words	English Manitoba	Immersion Manitoba	**2.06**	.025
		English Ottawa	**-3.33**	<.001
	Immersion Manitoba	English Ottawa	**-5.39**	<.001
Pre-algebra	English Manitoba	Immersion Manitoba	-.59	.586
		English Ottawa	**-2.53**	<.001
	Immersion Manitoba	English Ottawa	**-1.94**	<.001
Problem-solving	English Manitoba	Immersion Manitoba	-.48	.479
		English Ottawa	**-1.44**	.002
	Immersion Manitoba	English Ottawa	**-.97**	.027

Appendix C. Accuracy Rate of Math Words Measure by Grade.

In general, grade 3 students had higher accuracy rate in mathematical terms that were selected from lower grades (i.e., from grade 1 to grade 3), and accuracy rate was lower in items that were selected from higher grades (i.e., from grade 4 to grade 8).

Table 14. *Accuracy rate of Math Words Measure by Grade.*

	grade 1	grade 2	grade 3	grade 4	grade 5	grade 6	grade 7	grade 8
n	2	4	11	8	6	4	4	1
Imm-Manitoba	68.2	40.63	40.8	22.73	33.15	30.13	26.43	17
Eng-Manitoba	70.35	64.85	44.99	33.48	33.6	26.27	27.13	18.6
Eng-Ottawa	88.5	64.95	53.91	36.93	44.63	33.33	42.5	31

References

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2008). Evaluating the validity of the Automated Working Memory Assessment. *Educational Psychology*, *28*(7), 725–734. https://doi.org/10.1080/01443410802243828

Amen, J. (2006). *Using math vocabulary building to increase problem solving abilities in a 5th grade classroom*.

Ashcraft, M. H. (2002). Math Anxiety: Personal, Educational, and Cognitive Consequences. *Current Directions in Psychological Science*, *11*(5), 181–185. https://doi.org/10.1111/1467-8721.00196

Bae, Y. S., Chiang, H.-M., & Hickson, L. (2015). Mathematical Word Problem Solving Ability of Children with Autism Spectrum Disorder and their Typically Developing Peers. *Journal of Autism and Developmental Disorders*, *45*(7), 2200–2208. https://doi.org/10.1007/s10803-015-2387-8

Baker, C. K., & Galanti, T. M. (2017). Integrating STEM in elementary classrooms using model-eliciting activities: Responsive professional development for mathematics coaches and teachers. *International Journal of STEM Education*, *4*(1), 10. https://doi.org/10.1186/s40594-017-0066-3

Berg, D. H. (2008). Working memory and arithmetic calculation in children: The contributory roles of processing speed, short-term memory, and reading. *Journal of Experimental Child Psychology*, *99*(4), 288–308. https://doi.org/10.1016/j.jecp.2007.12.002

Brown, T. (1994). Creating and knowing mathematics through language and experience. *Educational Studies in Mathematics*, *27*(1), 79–100. https://doi.org/10.1007/BF01284529

Chan, W. W. L., & Wong, T. T.-Y. (2019). Subtypes of mathematical difficulties and their stability. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000383

Common Core State Standards Initiative. (2010). *Common Core State Standards for Mathematics.*

Connolly, A. J. (2007). *KeyMath-3 Diagnostic Assessment*. Pearson.

DeStefano, D., & LeFevre, J. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology*, *16*(3), 353–386. https://doi.org/10.1080/09541440244000328

Douglas, H. (2020). *Fraction symbols and their relation to conceptual fraction knowledge for students in grades 4 and 6.* Carleton University.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test Third Edition: Examiner's manual for the Ppvt-III Form IIIA and Form IIIB.* Circle Pines, MN: American Guidance Service.

English, L. D. (2017). Advancing Elementary and Middle School STEM Education. *International Journal of Science and Mathematics Education*, *15*(S1), 5–24. https://doi.org/10.1007/s10763-017-9802-x

Fetzer, M., & Tiedemann, K. (2018). The Interplay of Language and Objects in the Process of Abstracting. In J. N. Moschkovich, D. Wagner, A. Bose, J. Rodrigues Mendes, & M. Schütte (Eds.), *Language and Communication in Mathematics Education* (pp. 91–103). Springer International Publishing. https://doi.org/10.1007/978-3-319-75055-2_8

Fletcher, M., & Santoli, S. (2003). *Reading to learn concepts in mathematics: An action research project.* ERIC Document Reproductive Service No. ED 482 001. https://files.eric.ed.gov/fulltext/ED482001.pdf

Forsyth, S. R., & Powell, S. R. (2017). Differences in the Mathematics-Vocabulary Knowledge of Fifth-Grade Students With and Without Learning Difficulties: FIFTH-GRADE MATHEMATICS VOCABULARY. *Learning Disabilities Research & Practice*. https://doi.org/10.1111/ldrp.12144

Fuchs, L. S., DeSelms, J., & Deason, J. (2012). *Word problem-specific language assessment*. Vanderbilt University.

Fuchs, L. S., Fuchs, D., Compton, D. L., Hamlett, C. L., & Wang, A. Y. (2015). Is Word-Problem Solving a Form of Text Comprehension? *Scientific Studies of Reading*, *19*(3), 204–223. https://doi.org/10.1080/10888438.2015.1005745

Hambleton, R. K., Swaminatha, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.

Harmon, J. M., Hedrick, W. B., & Wood, K. D. (2005). Research on Vocabulary Instruction in the Content Areas: Implications for Struggling Readers. *Reading & Writing Quarterly*, *21*(3), 261–280. https://doi.org/10.1080/10573560590949377

Hughes, E. M., Powell, S. R., & Lee, J.-Y. (2020). Development and Psychometric Report of a Middle-School Mathematics Vocabulary Measure. *Assessment for Effective Intervention*, *45*(3), 226–234. https://doi.org/10.1177/1534508418820116

Kostos, K., & Shin, E. (2010). Using Math Journals to Enhance Second Graders' Communication of Mathematical Thinking. *Early Childhood Education Journal*, *38*(3), 223–231. https://doi.org/10.1007/s10643-010-0390-4

LeFevre, J.-A., Fast, L., Skwarchuk, S.-L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to Mathematics: Longitudinal Predictors of

Performance: Pathways to Mathematics. *Child Development*, *81*(6), 1753–1767. https://doi.org/10.1111/j.1467-8624.2010.01508.x

Lin, X., Peng, P., & Zeng, J. (2021). Understanding the Relation between Mathematics Vocabulary and Mathematics Performance: A Meta-analysis. *The Elementary School Journal*, *121*(3), 504–540. https://doi.org/10.1086/712504

Manitoba Education. (2013). *Kindergarten to Grade 8 Mathematics: Manitoba Framework of Outcomes.* Government of Manitoba.

Monroe, E. E., & Panchyshyn, R. (1995). Vocabulary Considerations for Teaching Mathematics. *Childhood Education*, *72*(2), 80–83. https://doi.org/10.1080/00094056.1996.10521849

Morsanyi, K., McCormack, T., & O'Mahony, E. (2018). The link between deductive reasoning and mathematics. *Thinking & Reasoning*, *24*(2), 234–257. https://doi.org/10.1080/13546783.2017.1384760

O'Halloran, K. (2005). *Mathematical discourse: Language, symbolism and visual images.* Continuum.

Ontario Ministry of Education. (2005). *The Ontario Curriculum, Grades 1-8: Mathematics.*

Peng, P., & Lin, X. (2019). The relation between mathematics vocabulary and mathematics performance among fourth graders. *Learning and Individual Differences*, *69*, 11–21. https://doi.org/10.1016/j.lindif.2018.11.006

Peng, P., Lin, X., Ünal, Z. E., Lee, K., Namkung, J., Chow, J., & Sales, A. (2020). Examining the mutual relations between language and mathematics: A meta-analysis. *Psychological Bulletin*, *146*(7), 595–634. https://doi.org/10.1037/bul0000231

Powell, S. R., Bos, S. E., & Lin, X. (2021). The assessment of mathematics vocabulary in the elementary and middle school grades. In A. Fritz, E. Gürsoy, & M. Herzog (Eds.), *Diversity Dimensions in Mathematics and Language Learning* (pp. 313–330). De Gruyter. https://doi.org/10.1515/9783110661941-016

Powell, S. R., Driver, M. K., Roberts, G., & Fall, A.-M. (2017). An analysis of the mathematics vocabulary knowledge of third- and fifth-grade students: Connections to general vocabulary and mathematics computation. *Learning and Individual Differences*, *57*, 22–32. https://doi.org/10.1016/j.lindif.2017.05.011

Powell, S. R., & Nelson, G. (2017). An Investigation of the Mathematics-Vocabulary Knowledge of First-Grade Students. *The Elementary School Journal*, *117*(4), 664–686. https://doi.org/10.1086/691604

Purpura, D. J., Hume, L. E., Sims, D. M., & Lonigan, C. J. (2011). Early literacy and early numeracy: The value of including early literacy skills in the prediction of numeracy development. *Journal of Experimental Child Psychology*, *110*(4), 647–658. https://doi.org/10.1016/j.jecp.2011.07.004

Purpura, D. J., & Logan, J. A. R. (2015). The nonlinear relations of the approximate number system and mathematical language to early mathematics development. *Developmental Psychology*, *51*(12), 1717–1724. https://doi.org/10.1037/dev0000055

Purpura, D. J., Logan, J. A. R., Hassinger-Das, B., & Napoli, A. R. (2017). Why do early mathematics skills predict later reading? The role of mathematical language. *Developmental Psychology*, *53*(9), 1633–1642. https://doi.org/10.1037/dev0000375

Purpura, D. J., & Reid, E. E. (2016). Mathematics and language: Individual and group differences in mathematical language skills in young children. *Early Childhood Research Quarterly*, *36*, 259–268. https://doi.org/10.1016/j.ecresq.2015.12.020

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, *20*(2), 110–122. https://doi.org/10.1016/j.lindif.2009.10.005

Riccomini, P. J., Smith, G. W., Hughes, E. M., & Fries, K. M. (2015). The Language of Mathematics: The Importance of Teaching and Learning Mathematical Vocabulary. *Reading & Writing Quarterly*, *31*(3), 235–252. https://doi.org/10.1080/10573569.2015.1030995

Rubenstein, R. N., & Thompson, D. R. (2002). Understanding and Supporting Children's Mathematical Vocabulary Development. *Teaching Children Mathematics*, *9*(2), 107–112. https://doi.org/10.5951/TCM.9.2.0107

Sahin, A., & Anil, D. (2017). The Effects of Test Length and Sample Size on Item Parameters in Item Response Theory. *Educational Sciences: Theory & Practice*. https://doi.org/10.12738/estp.2017.1.0270

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

Song, C. S., Xu, C., Maloney, E. A., Skwarchuk, S.-L., Di Lonardo Burr, S., Lafay, A., Wylie, J., Osana, H. P., Douglas, H., & LeFevre, J.-A. (2021). Longitudinal relations between young students' feelings about mathematics and arithmetic performance. *Cognitive Development*, *59*, 101078. https://doi.org/10.1016/j.cogdev.2021.101078

StataCorp. (2017). *Stata Statistical Software: Release 15.* StataCorp LLC.

Toll, S. W. M., & Van Luit, J. E. H. (2014). The Developmental Relationship Between Language and Low Early Numeracy Skills Throughout Kindergarten. *Exceptional Children*, *81*(1), 64–78. https://doi.org/10.1177/0014402914532233

Ünal, Z. E., Powell, S. R., Özel, S., Scofield, J. E., & Geary, D. C. (2021). Mathematics vocabulary differentially predicts mathematics achievement in eighth grade higher- versus lower- achieving students: Comparisons across two countries. *Learning and Individual Differences*, *92*, 102061. https://doi.org/10.1016/j.lindif.2021.102061

Vygotsky, L. S. (1986). *Thought and language.* MIT Press.

Wechsler, D. (2014). *WISC-V: Technical and interpretive manual.* NCS Pearson Inc.

Xu, C., Lafay, A., Douglas, H., Di Lonardo Burr, S., LeFevre, J.-A., Osana, H. P., Skwarchuk, S.-L., Wylie, J., Simms, V., & Maloney, E. A. (2022). The role of mathematical language skills in arithmetic fluency and word-problem solving for first- and second-language learners. *Journal of Educational Psychology*, *114*(3), 513–539. https://doi.org/10.1037/edu0000673