Summer 7-28-2016

Evaluate the expression: An analysis of the linguistic features of directions in an upper elementary mathematics textbook

Melinda Freeland
Hamline University, mfreeland01@hamline.edu

EVALUATE THE EXPRESSION: AN ANALYSIS OF THE LINGUISTIC FEATURES OF DIRECTIONS IN AN UPPER ELEMENTARY MATHEMATICS TEXTBOOK

by

Melinda LeFeber Freeland

A capstone submitted in partial fulfillment of the requirements for
the degree of Master of Arts in English as a Second Language.

Hamline University

Saint Paul, Minnesota

July 2016

Primary Advisor: Anne DeMuth
Secondary Advisor: Kristin Liu
Peer Reader: Olivia Disselkamp
Peer Reader: Rebecca Streeter

To my husband, my constant source of support, encouragement, and food.
You are the one.
And to the incredible team
who helped me every step of the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

A child raises her hand during a math test.

"Teacher, what is this asking me to do?"

I encourage her to read it out loud.

"Graph the two coordinate points on the grid below (1 point each)."

I think back to having this child walk out coordinate points on a taped grid on the floor and encourage her to do her best. An upper level English learner (EL), she will hopefully remember the different vocabulary terms I taught throughout the past weeks.

"But teacher, how many points are there? What do I have to put on the graph?" the quiet voice persists as the child indicates with a finger both mentions of *points* and *point* in the prompt.

This instance of the language used in directions causing confusion is more common than not. While mathematics instructors in elementary classrooms will often include some vocabulary instruction, often there are other linguistic features used by the instructor that are assumed understood by the students. What are the specific aspects of the language used in a mathematics classroom that necessitate instruction? How do academic, content, and everyday vocabulary intersect to create a unique language in mathematics? This paper will analyze the linguistic features of the directional language of mathematics in a mainstream fourth grade mathematics textbook in order to identify features that should be explicitly taught to English learners (ELs), with the goal of making teachers more aware, allowing them to better support their ELs in acquiring mathematics language and content. This chapter provides a preface for studying the language of mathematics. It also shares real-life situations that ground this research in the overarching need of providing explicit instruction around the language of mathematics for language learners.

The Language of Mathematics

The teaching and learning of mathematics is often seen as a process devoid of language; many mathematics teachers believe that mathematical concepts are universal and therefore mathematics learning is less language dependent than other content areas (Lager, 2006; Zevenbergen, 2000). Learning another language, however, involves more than learning the individual vocabulary words and the syntax of communication (Moschkovich, 2012). In order to achieve proficiency in any language, including the language of mathematics, learners not only need to understand that language, but be able to communicate with it and be understood.

The field of mathematics is seen by some researchers and instructors as possessing a distinct language of communication, with oral, written, formal, and informal aspects (Adams, 2003; Adler, 1998; Barwell, 2005; Irujo, 2007; Joseph, 2012; Kessler, Quinn, & Hayes, 1985; Lager, 2006; Moschkovich, 2012; de Oliveira & Cheng, 2011; Temple & Doerr, 2012; Usiskin, 1996; Zevenbergen, 2000; Zwiers, 2008). The language of mathematics can also be considered an academic aspect of any language, as it is not commonly spoken at home and is typically learned at school. Like other languages, the language of mathematics is expressed orally, in written form, and is informal or formal with a focus on communicating meaning. It possesses internal logic and relationships between words and structure unique to the field of mathematics (Schleppegrell, 2007; Usiskin, 1996). Writing takes a preeminent role as the distinct symbols in mathematics are more easily conveyed in written form than orally. There are also words that have unique meanings in the mathematics register, such as *if* and *random*.

While it is important to interact orally with the language of mathematics in order to understand and internalize it, the spoken language can be much more abstract and often removed from its visual or pictorial representation (Francis, Rivera, Lesaux, Kieffer, & Rivera, 2006; Schleppegrell, 2007; Usiskin, 1996). In studies of the field of mathematics, researchers have found terms specific to the field are often explicitly taught, but the other academic language features are identified as having less frequent and systematic instruction (Haag, Heppt, Stanat, Kuhl, & Pant, 2013; Lager, 2006). For example, when a student reads the direction "Express the following fractions in simplest form," it would be assumed that the terms *fraction* and *simplest form* would have been taught in context of learning the content. But the terms *express* and *following* might have been used in communication without being explicitly defined. Students must learn all components of the language of mathematics if they are to be effective speakers, listeners, and communicators in the mathematics classroom.

Although learning a language involves learning through language and learning about language, while also going through the stages of language acquisition (Achugar, Schleppegrell, & Oteíza, 2007), researchers differ over the need for explicit teaching of academic and content terms used in the mathematical context. Explicit teaching could potentially distract from mathematics instruction, but relying on experiences where students hear and use mathematical language might not push them to move beyond ordinary discourse. This tension of explicit versus implicit instruction becomes more complicated as the language of mathematics contains technical vocabulary in dense noun phrases, clauses with verbs of being, and complex sentences with specialized conjunctions, creating unique lexical bundles (Temple & Doerr, 2012). The way that language is utilized by the teacher can add another layer of complication, heightening the need to provide explicit language instruction.

Within a focus on language instruction, building academic language is important for students' success in school. Learning academic language is more than just learning a list of terms; it is developing skills through which students can make meaning (Moschkovich, 2012). Features of acquiring academic language include pronunciation and intonation, identifying and knowing which terms to use and which to exclude, understanding particular meanings of words, utilizing preferred sentence structures and accepted discourse patterns, and the pragmatic rules and use of language to accomplish a task (Irujo, 2007); in addition, students need to learn "mortar terms" that would typically lack a tangible definition. These words provide the connections or transitions around the academic terms in order to convey a complete thought (Zwiers, 2008).

Being able to participate fully in a content area requires proficiency in language specific to that content that is context-reduced and cognitively demanding; this is a more difficult aspect of language learning (Kessler et al., 1985; Francis et al., 2006; Lager, 2006; Zwiers, 2008). If ELs have previous schooling and therefore academic proficiency in their first language in the area of mathematics, this can be a foundation to develop such proficiency in their second language. If this is not the case, students will have to learn not only a second language, but also the cognitive, academic language of mathematics communicated through the second language (Dale & Cuevas, 1992). All of these aspects and layers of the language of mathematics, as well as the need for explicit instruction on more than basic content vocabulary, will be taken into consideration in the analysis of the directional language of the fourth grade *Math in Focus* mathematics textbook to identify linguistic features that should be explicitly taught.

Role of the Researcher

The curriculum chosen for this research, *Math in Focus*: *Singapore Math by Marshall Cavendish* (Fong, Ramakrishnan, & Gan, 2009) was selected because of my intimate knowledge of the curriculum and six years of experience teaching with the fourth and fifth grade materials. While I have also taught *Saxon Math* and *Everyday Mathematics*, the two schools that implemented *Math in Focus* chose this curriculum specifically for its rigor and higher level language, especially in word problems (Fong, 2009). Originally developed in Singapore as the *My Pals Are Here! Maths* curriculum, the goal is to help children master mathematics concepts, computational skills, problem-solving skills, and apply mathematics activities in daily life by promoting creative, critical, and inquiry-based thinking (Fong, 2009).

While there are pros and cons to the use of the *Math in Focus* curriculum because of its problem-solving demands, the aspect relevant to this research study is unaffected by them; the directional language in the fourth grade textbook remains a constant in the instructional setting, unlike the changes that can occur based on how an instructor presents and utilizes the curriculum. The benefit of identifying linguistic features in the instructional text, as well as the use of everyday terms and phrases, will allow any teacher to recognize which terms are prevalent in a mathematics curriculum and therefore necessitate instruction. It also removes any personal variations of how instructors communicate directions in the mathematics classroom, focusing on the foundation of mathematical language present in every classroom that uses the *Math in Focus* texts. Because of that, in this study the role of the researcher is to collect, analyze, and interpret the linguistic data, making the resulting information available for instructional purpose in the language and mathematics classrooms.

Background of the Researcher

The context of mathematics was not chosen randomly. Throughout my years as an upper elementary classroom teacher, I have always had a moderate to high percentage (39%-99%) of ELs in every class. Those language learners were in my classroom for the majority, if not all, of their instruction, making it an inclusive classroom setting. I initially assumed that the mathematics content would be less language dependent and therefore more easily learned than the language arts content. This was quickly proven false, as interacting with various ELs led me to uncover several language and instructional features unique to the field of mathematics. I began exploring other means of supporting mathematics instruction for language learners, such as pursuing a license in English as a second language (ESL). The more information I gathered, the more I realized that ELs would benefit from focused mathematics instruction, specifically in regards to developing the language of mathematics. That learning also came from my experiences working with ELs who exemplified the idea that language and content instruction both need to occur in the study of mathematics. It was through the experiences of assessing their background knowledge and adjusting instruction and activities to better meet their needs that I came to recognize that the field of mathematics possesses its own language; in order to better instruct ELs, I would need an additional knowledge base and set of resources.

Theresa was a Karen immigrant who came from a refugee camp in Thailand and the next day started her first day of school in my mainstream, fifth grade classroom. (This name and all subsequent ones are pseudonyms.) At the same time, Nan Dah, another Karen refugee, entered my classroom. She came from a different camp in Thailand where she had about a year of basic English education in math and reading. Her parents were college graduates and taught Nan Dah to read and write in Thai and Karen. With no resources and EL pull-out only to teach them common school and community vocabulary, I quickly realized that in order to better serve these students I would need to create a supplementary math program that would allow them to access the grade-level curriculum. Using basic terms, manipulatives, and real-world objects, I was able to mimic and create enough patterns that Nan Dah was able to connect to the simple adding, subtracting, and multiplying she had learned in the camp. She was then able to build on her knowledge by recognizing numbers in other forms, while also learning the vocabulary around the concepts she already knew. In contrast, removing the language and providing manipulatives quickly proved not to be enough for Theresa, as she had no concept of numbers representing quantities in English or in Thai. While throughout the year there proved to be other developmental hindrances to Theresa’s progress, this was the first time I realized the majority of the mathematics curriculum taught in the upper elementary relies on prerequisite skills and a solid concept of number sense, both of which necessitate comprehension of the language of mathematics.

Another case that further prompted my search for resources in teaching upper elementary mathematics to ELs, or more specifically to students with limited or interrupted formal education (SLIFE) was Celeste. An immigrant from Mexico, she entered fifth grade at an early second grade level in reading and writing in Spanish. For a newcomer, her social English quickly proved to be higher than the average. After observing her struggle to apply a mathematics concept from one day to the next, such as multiplying by tens, I created a couple "naked number" problems to identify if Celeste was able to perform basic computation. These problems utilized only digits and computational symbols, with a few solved to illustrate patterns. In attempting to solve the problems, Celeste demonstrated she was not able to regroup in addition and subtraction, let alone understand the concepts behind multiplication and division; she was also unable to count past the low hundreds in Spanish and did not know place value around money or numbers past the tens place. Entering a higher grade level as a SLIFE not only means that Celeste had to learn a new language, but also learn what it means to participate in school and learn. Because Celeste started halfway through the school year when we were learning how to multiply fractions, the gap between what she knew and what we were trying to work with only widened. Putting her on a language-removed, visual computer program designed by the Mind Research Institute to teach basic math skills in a visual, conceptual manner only proved to frustrate her, as the visual representations of numbers did not relate to any real-world experience she possessed with amounts and spatial concepts. She also struggled to connect how moving the ten buttons from one box to the next represented regrouping in a base-ten number system. Providing instruction in Spanish was also not a viable resource, as most of the mathematics concepts were completely new to her; she was learning new academic vocabulary in both languages instead of one. While the attempted solutions above only raised more problems, Celeste was able to make slow progress throughout the year in basic computation, while remaining far below grade level.

A final example of a student who prompted me to not only pursue other approaches and resources to teaching mathematics, but also to more specifically focus on language in the mathematics classroom, was Oliver. With five years of previous schooling in Mexico, Oliver quickly demonstrated grade-level proficiency in reading and writing in his native language. Because of my work with the previous students, and specifically Celeste, I had already created basic mathematics assessments that were language removed and built up to the skills needed to succeed in fifth grade mathematics. After several formative assessments, I determined that Oliver was at grade level in mathematics; while his previous instruction was in Spanish, most of the concepts seemed familiar to him. He also possessed a strong ability to recognize and build on patterns, a critical thinking skill which allowed him to quickly progress in all areas of learning, for both content and language. While he carried over the skills knowledge in mathematics, and I explicitly focused on vocabulary at the beginning of each new unit, he still struggled with some of the language of the directions and multi-step word problems. I quickly realized Oliver did not struggle on the terms that we would typically assign to mathematics, such as *circumference* or *division*, but instead was confused by seemingly common terms used in an unfamiliar way, such as *identify the point* or *the measure is about 100 meters*. While he would be able to define each term separately, when they were used together to convey a concept in mathematics, he struggled. I resorted to trying to explain the terms as situations of confusion arose, but continued to feel frustrated that I was not teaching Oliver the language he needed before he encountered the directions or problems.

Each of these stories identifies how I was further pulled down the path of the language of mathematics. Throughout my interactions with these students, I struggled to find resources or know where to begin providing instruction around the various types of academic and everyday vocabulary used in the field of mathematics. Each of these students was required to interact with, and ultimately be tested on, the mathematics content provided in the *Math in Focus* textbook; one clear area to explore that might meet this identified need would be to analyze the language used in the text in order to identify a direction for instruction. My background in teaching this curriculum, as well as providing training for colleagues new to this curriculum, gives me additional familiarity in how to identify and integrate steps for language instruction in the mathematics classroom. Analyzing linguistic features of the directions in a mathematics textbook may seem separate from the day-to-day struggles in teaching language learners English as well as mathematics content. But the implications of identifying features in the language of mathematics that are typically ignored could prove beneficial in better preparing learners to acquire the domain-specific language.

Guiding Questions

My experiences in various upper elementary mathematics classrooms have led me to several questions. First, what are common features of the language of mathematics? How do academic, content, and everyday vocabulary intersect to create a unique language in the mathematics content? What features of the language of mathematics are commonly used in directions in textbooks? More specifically, what are the linguistic features of the language of mathematics that are difficult to understand in directions or instructions? Are there features present in directions that are not found in the basic content or recommended vocabulary that is typically taught? Finding the answers to these questions starts with analyzing the linguistic features in the directional sentences in an upper elementary mathematics textbook in order to identify common features that should be explicitly taught to ELs; this identification will further support ELs in achieving proficiency in the language of mathematics.

Summary

In this introduction, I provided an overview of the features of the language of mathematics that make it a unique language, and the circumstances encountered in upper elementary inclusive classrooms where ELs demonstrate a language need that is not currently met by available resources, language learning strategies, or instructional systems. There are language features in use that, if taught, would better support ELs in their language acquisition; specifically in regards to a mathematics curriculum, instruction would support not only their acquisition of the language of mathematics, but also the content. Because of this, it is important to identify the linguistic features unique to the field of mathematics.

Chapter Overviews

In Chapter One the topic was introduced by identifying the need and significance of this research. The context of the study was introduced, as well as the background and role of the researcher. Chapter Two provides a review of literature relevant to the linguistic features of the language of mathematics, the specific needs of ELs, and the benefit of explicit language instruction. Chapter Three describes the research design and methodology used in the subsequent curriculum analysis, while Chapter Four presents the results. In Chapter Five there is a reflection on the data collected and identification of next steps to utilize this information in supporting ELs' acquisition of the language of mathematics. The final chapter also discusses limitations of this study and recommendations for further research.

CHAPTER TWO: LITERATURE REVIEW

This paper analyzes the linguistic features of the directional language of mathematics in a mainstream fourth grade mathematics textbook in order to identify features that should be explicitly taught to ELs. It also looks at the vocabulary or language recommended for instruction compared to that which is used in a mathematics textbook. The focus is to identify the common features of the language of mathematics and learn how academic, content, and everyday vocabulary intersect to create a unique language in the mathematics content. The goal of the literature review is to answer the following questions. First, what are common features of the language of mathematics? How do academic, content, and everyday vocabulary intersect to create a unique language in the mathematics content? What features of the language of mathematics are commonly used in directions in textbooks? More specifically, what are the linguistic features of the language of mathematics that make it difficult to comprehend directions or instructions? Are there features present in directions that are not found in the basic content or recommended vocabulary that is typically taught?

The mathematics register provides the context in which students need to be proficient in the language of mathematics in order to understand and be understood (Zwiers, 2008). In the field of mathematics, there are linguistic structures that are used differently than how those same structures are applied in everyday life. Researchers define the mathematics register as the meanings conveyed through words in the language of mathematics, where students need to learn new ways to utilize and combine these styles to communicate meaning (Schleppegrell, 2007). This language specific to mathematics contains many different aspects, including symbols, oral language, written language, and visual representations. Students need to learn the notational forms created by mathematicians as well as the English words that identify them for use in mathematical discourse (Spanos, Rhodes, Corasaniti, & Crandall, 2013). Even the position, order, orientation, or size of features of the text conveys meaning in mathematics, as this symbolism conveys relationships and patterns in ways that everyday language cannot (Schleppegrell, 2007; Zwiers, 2008).

The difficulties in acquiring the language of mathematics can be traced to the syntactic, semantic, and pragmatic features that must be understood in order for students to verbalize or interpret mathematical rules and concepts in English. Understanding how these linguistic features work to make meaning is integral to learning mathematics. While symbols and visual representations are part of the mathematical language, the grammatical structure of mathematics is a feature that might receive less attention in the classroom, as might the multiple layers of meaning present in the vocabulary and discourse of mathematics. Martiniello (2008) found in her study of the linguistic features of mathematical word problems that overall reading comprehension includes reading fluency, vocabulary and syntactic knowledge, and discourse comprehension. Comprehending mathematical text is difficult because of the added aspect of text combined with symbols and differing forms of orientation in layout.

The purpose of Chapter Two is to review recent research on the language of mathematics in order to identify the features that might make understanding written directions difficult for ELs. First the syntax, semantics, and discourse patterns of the language of mathematics are described as identified in studies around the linguistic features of mathematics. Then the specific needs of ELs in learning the language of mathematics are connected to the complexity of the mathematics register. This research review will help identify the common features of the language of mathematics that are difficult to comprehend, as well as explore how academic, content, and everyday language intersect to create a unique language in the mathematics register.

The Language of Mathematics

Research on Mathematics Assessments and Textbooks

The linguistic features common to the language of mathematics described throughout this review have been identified through various research studies that primarily focus on mathematics assessments and word problems. Six studies analyze the discrepancy between ELs performance on comprehensive mathematics assessment compared to English speaking peers, identifying linguistic features in word problems within these assessments that might be impacting ELs performance (Abedi & Lord, 2001; Bergqvist, Dyrvold, & Österholm, 2012; Haag et al., 2013; Lager, 2006; Martiniello, 2008; Shaftel, Belton-Kocher, Glassnapp, & Poggio, 2006). One of these studies, conducted by Bergqvist et al. (2012), created a corpus from two eighth grade mathematic textbooks, representing mathematical language, and a corpus of everyday language from 58 current novels and newspapers. The words present in these corpora were then compared to the mathematical tasks in a comprehensive assessment, analyzing the amount of each language type and other linguistic features found in the state assessment. Bergqvist et al.

(2012) then looked at student performance, finding a correlation between ELs underperformance compared to peers and the presence of language aspects that influence task difficulty.

Analyzing the linguistic aspects of mathematics assessments and textbooks is continued by various other researchers. Those researchers also identify linguistic aspects that lead to intricacy within a mathematics text, whether they compare the development of linguistic complexity across levels (Monaghan, 1999), categorize the language used to convey writing tasks (Joseph, 2012), or analyze the linguistic features common in word problems (Butler, Bailey, Stevens, & Huang, 2004a; Butler, Lord, Stevens, Borrego, & Bailey, 2004b; Sweeney, 2014). There is not currently research around the language used for instruction or directions in mathematics textbooks; as this is most commonly the section students encounter without a teacher's guidance or direct instruction, considered the guided and independent practice sections of the lesson, this review identifies a need for addressing this specific topic. Supporting ELs with navigating a mathematics text is essential, as the language of mathematics contains several linguistic features that affect an ELs' ability to acquire and apply mathematical understanding (Abedi & Lord, 2001; Bergqvist et al., 2012; Butler et al., 2004a; Butler et al., 2004b; Haag et al., 2013; Joseph, 2012; Lager, 2006; Martiniello, 2008; Monaghan, 1999; Shaftel et al., 2006; Sweeney, 2014).

Syntax

Part of the density of the mathematics language is created by the use of complex structures in sentence construction. One of these structures is a noun phrase used to convey an abstract concept of the subject (*the volume of*), classify the subject (*right* angle), or qualify it (a number *which is a multiple of 3*). Mathematical operations are structured as things in noun phrases instead of processes, and students need to be able to recognize the relationship of things in grammar with the required processes of mathematical thinking (Schleppegrell, 2007). For example, the phrase: *The sum of two prime numbers is 8*, uses *sum* as a noun, while to actually identify the prime numbers students need to view *sum* as a process. Relational processes conveyed through phrases linked by *be* and *have* are another unique grammatical feature of the language of mathematics. These clauses are used to attribute membership to a class or relationship (*a rectangle is a quadrilateral*) or identify, describe, or equate a relationship (*an even number is a number that can be equally divided by two* or *a fraction has a numerator and a denominator*). These relational constructions can be difficult; not only might students be accustomed to conveying relationships in different ways, but the point of view is not readily apparent (e.g., the difference between the properties of a figure versus the categories of classification) (Schleppegrell, 2007). In addition, passive verbs can hide the doer of an action, causing readers to be unsure of what action is taking place and who is completing it (Zwiers, 2008). Martiniello (2008) has found in her study of the linguistic features of word problems in a mathematics assessment that long noun phrases, lack of clear relationship between syntactic units, and multiple clauses were the syntactical features that hindered ELs' mathematical performance. These three aspects of syntax work together to convey information that is often complicated and abstract in the language of mathematics.

Noun phrases. The term *noun phrase* refers to a noun and its accompanying modifiers. The linguistic load of a noun phrase is complicated by the multiple elements that can be a part of a noun phrase, including pronouns, referents, relative clauses, and negation. When the words are taken separately, students might be able to identify most of them; used together the words often convey a unique or more complicated meaning, increasing the processing burden for students (Butler et al., 2004a). In an analysis of the discourse of mathematics, de Oliveira and Cheng (2011) found that noun phrases in mathematical text present information as precise, authoritative and technical. More frequently than not, the noun phrases they analyzed were long and complex, combining a head noun with numerous pre- and post-modifiers. The modifications communicated specific requirements or aspects of a task that needed to be solved, but often the density of the nominal group obscured the overall meaning (Butler et al., 2004a; de Oliveira & Cheng, 2011). For example: “Multiply the value of the digits in the greatest place of each number” (Fong et al., 2009, 4A p. 36). The head noun, *value*, is obscured by the various elements as they continue to build a layer descriptions of what specific parts to multiply.

The use of pronouns also adds to the difficulty of mathematical texts, as the reference to another sentence element might cause ambiguity for the language learner (Shaftel et al., 2006). The following problem contains several different pronouns that could lead to confusion: “Novak bought a box of 72 building blocks. He shared the blocks equally with his 2 friends. How many blocks did each of them get?” (Fong et al., 2009, 4A p. 37). The pronoun *them* could cause confusion as to whether it refers to just Novak’s friends or if it also includes him. A misunderstanding with this one term would lead to an error in the final answer, let alone other complexities with the language used that could cause confusion.

Students need to learn to identify how language repackages the process of mathematics reasoning, particularly when long noun phrases are employed, as in the following problem: "If a rectangular solid has side, front, and bottom faces with areas of 2x, y/2, and xy cm^2 respectively, what is the volume of the solid in centimeters cubed?" (Fang & Schleppegrell, 2010, p. 590). The head noun, *faces*, not only has pre-modifiers, *side, front,* and *bottom,* but also post-modifiers: *with areas of 2x, y/2, and xy* cm^2 *respectively*. This requires the application of several mathematical processes to solve, as well as an understanding of the effect of the verb in conveying the connection (Fang & Schleppegrell, 2010). The use of numbers as nouns can also be a confusing structure as the quantity is taken as a whole and referred to as one unit or thing (Dale & Cuevas, 1992). When numbers are used as modifiers, such as *two cubes*, being able to recognize and utilize this information becomes another level of mathematical interpretation.

Another intricacy within a noun phrase occurs with the use of referents. Part of a learner's ability to decode complex noun phrases involves knowing how reference is indicated. Problems using referents such as *the number* or *a number* require the reader to infer what is being referenced. For example: "When 15 is added to a number, the result is 12. Find the number" (Dale & Cuevas, 1992, p. 334). In order to solve this, the reader must figure out that *a number* and *the number* refer to the same unknown quantity, and that *the number* can be expressed symbolically in terms of the two other numbers given.

Along with referents, relative clauses are an aspect of nominal clauses that qualify the head noun. The information they provide is either needed to determine the item being referenced, and therefore is considered a defining relative clause, or the information is additional and not essential, and therefore is a non-defining relative clause (Derewianka, 2013). Relative clauses, defining or non, can obscure the composition of the clause and therefore be difficult for ELs to understand what is being described (Abedi & Lord, 2001). For example, the following sentence contains two relative clauses: “The number of medals won by top ranking countries in the 2006 Winter Olympics held in Turin, Italy is recorded in the table” (Fong et al., 2009, 4A p. 135). The base sentence is: *The number of medals is recorded in the table*, but in order to determine which *medals* are being referred to, the subsequent relative clauses provide the information needed to identify them, *won by top ranking countries*. The prepositional phrase *in the 2006 Winter Olympics* also has a relative clause, *held in Turin, Italy*, which gives more information to the *Winter Olympics*. These multiple relative clauses require students to track what is being referred to throughout the sentence, a process which might prove even more difficult for ELs.

Another convolution of noun phrases is the use of negation. Several research studies indicate that sentences containing negations are harder to comprehend than sentences phrased in the affirmative (de Oliveira & Cheng, 2011). Some languages, such as Spanish, retain the negation even when a double negative is employed, therefore adding to the lack of comprehension around negated statements for some language learners (Mestre, 2013). Removing negation is one way to increase comprehensibility, but it might lead to a decrease in complexity of problem solving. An example of this is the following direction from the fourth grade *Math in Focus* (Fong et al., 2009): “Explain which of the answers are unreasonable” (4A p. 107). Asking for the answers that are *unreasonable* increases the cognitive load, but the prefix indicating the opposite amount might go unnoticed and lead to an unknown mistake.

Verb Phrases. Verb phrases refer to the words that convey processes, whether it is doing, thinking, saying, or relating. A verb phrase can consist of one verb or many verbs, including auxiliaries, modals, prepositions, negations, and adverbs (Derewianka, 2013). The language of mathematics is considered complex because of its use of various verb forms, such as verb tenses, modals, and passive verb construction (Gerofsky, 1996). In word problems, the verb tenses are often combined in a way that would typically be deemed contradictory compared to language used in other mathematical contexts; verbs can move from present to past and again to present tense within a situation and therefore exemplify the struggle with word problems in that they do not refer to real places, people, or situations. Through the use of a variety of verb forms, word problems create a hypothetical situation, pretending that a particular story situation exists under the specific situation and direction of the author of the problem. If a student is unaware of this view and does not "buy in" to the world the word problem creates, she or he might be troubled by trying to answer irrelevant questions. For example, in the following problem, verb tenses are combined and a hypothetical world created in which the reader is able to predict what will happen in the future: "Each elephant at the Young Elephant Training Centre in Pang-ha, Thailand eats about 250 kg of vegetation in a day. How much would 43 elephants eat in 1 day? 1 week?" (Gerofsky, 1996, p. 40). The modal *would* requires the student to believe a world in which she or he can determine future events by expecting a mathematical pattern to continue in a hypothetical situation, where by changing the question to *How much do 43 elephants eat?*, the student would still need to believe the premise of the word problem, but not rely on conditional circumstances.

The way various verb forms are used can also prove difficult to understand. Passive verbs in mathematical questions often convey an action done to a number or figure in order to establish connection or naming (e.g., *is represented by*, *is read*, *have received*) (Irujo, 2007). Combined with nominalization, this passive structure often covers up agency and presents information as given and not actively engaged (de Oliveira & Cheng, 2011; Zwiers, 2008). For example: "If the answer is wrong, the cards are taken away from the player" (Fong et al., 2009, 4B p. 46). The passive structure describing the cards makes the information seem given, but instead the students who read these directions need to take an active role in carrying out the step. Students might have trouble identifying not only what is described as being accomplished in the passive phrase, but also who or what is involved (de Oliveira & Cheng, 2011). Sentences that begin with an imperative, such as *find, name,* or *evaluate,* address the reader as the performer of the action. The imperative implies that students are inducted into the mathematics community and elicits them to complete the action (de Oliveira & Cheng, 2011). Verbs, modals, and auxiliaries can be combined in phrases that include three or more words to create complex verbs (Shaftel et al., 2006). Such combinations also indicate that multiple or difficult verb tenses are employed, as *had been going* or *would have eaten* indicate past perfect progressive and future perfect with a conditional, respectively.

Structures impacting sentence complexity. The language of mathematics may also employ familiar structures, like prepositions, conjunctions, and logical connectors, in ways that make sentence structure more complex. Prepositions and conjunctions are used in a more precise and technical way in mathematics than in everyday communication (Butler et al., 2004a). With any type of multi-step reasoning, in word problems or theorems and proofs, words like *when*, *therefore*, *if*, and *given* convey relationships between clauses where the information in one is needed to complete or solve the other. In complex noun phrases, the post-modifiers are often embedded clauses, increasing the density of the nominal structure (Butler et al., 2004a; de Oliveira & Cheng, 2011). This relationship of information can even be conveyed subtly as researchers have found that mathematical texts often contain long patterns of reasoning without clearly indicating what operational properties, axioms, definitions or laws were assumed (Schleppegrell, 2007). Researchers have also demonstrated how mathematical discourse combines phrases and clauses to produce complex sentences, such as: “Which of the following is the *best* estimate for the total number of student speeches that could be given in a 2-hour class?” (Irujo, 2007, p. 4).

Prepositions alone can change the meaning of a sentence. Prepositions mark the start of an additional phrase and therefore an added concept to understand (Butler et al., 2004a; Shaftel et al., 2006). For example, *5 divided by 10 is ½* is very different from *5 divided into 10 is 2*, although only the preposition is changed (Irujo, 2007). Another syntactic pattern that often proves problematic is the combined prepositional phrase and passive voice, such as: *four divided into nine equals nine-fourths, nine divided by four equals nine-fourths, if nine is divided by four, nine-fourths results* (Spanos et al., 2013). This requires the student to not only understand what is conveyed in each part of the sentence, but also the meaning communicated through the passage as a whole.

Comparative structures play an important role in the field of mathematics, particularly since mathematics can be a study of relationships (Dale & Cuevas, 1992). Some of the confusion around clauses in the language of mathematics results from their use as indicators of comparison. Comparative structures such as *greater than, less than, n times as much as*, and *as large as* are confusing because while they possess specific meanings, their patterns can be used in a variety of ways (Butler et al., 2004b; Dale & Cuevas, 1992; Spanos et al., 2013). This means that students need to not only understand how these terms are used, but also select the meaning that is appropriate in the context. An example would be the difference between using *greater than* to compare *6 is greater than 3*, which can be represented by the symbol >, versus the question *What number is 3 greater than 6*, which asks students to add on in order to identify the specific greater number.

Logical connectors are words or phrases that mark a relationship between two or more clauses, such as similarity, contradiction, cause and effect, or chronological/logical sequence (Butler et al., 2004b; Dale & Cuevas, 1992). These connectors, which connect information by communicating the relationship between two concepts, require a language learner to understand both separately as well as their relationship together. *If* clauses alone possess multiple meanings, indicating causality, probability, or a change in result. The use of a logical connector combines semantics and syntax, as the structure of a connection affects the meaning (Irujo, 2007). *If* clauses in mathematics also indicate a hypothetical conditional, whereas in everyday language conjunctions such as *if*, *when*, and *so* are more vague and imprecise (Fang & Schleppegrell, 2010). Students must not only recognize logical connectors, but also the situations in which they appear and the meaning of their position within a sentence (Butler et al., 2004b; Dale & Cuevas, 1992). Again, this combination of understanding the term as well as its use in the context illustrates the syntactical difficulty of the language of mathematics.

Semantics

The language of mathematics combines multiple categories of vocabulary with symbols to convey meaning; students need proficiency in the linguistic features used in the field of mathematics. Kessler et al. (1985) argued that knowing vocabulary is more important than understanding English syntax and morphology. The language of mathematics has more distinct vocabulary terms that do not overlap with other academic domains, as well as vocabulary that changes depending on the linguistic structure (Rubenstein & Thompson, 2002; Zwiers, 2008). Rubenstein and Thompson (2002) identify some common challenges of vocabulary in the language of mathematics that highlight how the meaning changes (Appendix A). Looking at the complexity of everyday words combined with academic words, Rubenstein and Thompson demonstrate how the difference in meaning can cause confusion, such as *right angle* versus *right answer* versus *right hand*. Another layer of convolution lies in academic words that have different meanings in different contexts; divide in mathematics means to separate into parts, while the *Continental Divide* is the geographic line that marks the separation of water that flows east from water that flows west (Rubenstein & Thompson, 2002, p. 108). Understanding the nuances of meaning becomes more complicated as the different layers of language interact.

Vocabulary. Vocabulary in the mathematics language covers more than terms unique to the numbers and systems in mathematics. Knowing a word also involves recognizing and understanding its many uses (Dale & Cuevas, 1992). Vocabulary may be the most obvious linguistic feature in the mathematics register to analyze for indicators of linguistic difficulty, but recently research has explored how the mathematical syntax affects vocabulary. The language of mathematics uses everyday words, some possessing unique mathematical meanings, with conceptually dense academic and content specific terms, all structural aspects that impact the linguistic complexity (Butler et al., 2004a; Irujo, 2007; de Oliveira & Cheng, 2011). When terms from content, academic and everyday vocabulary combine to create intricate strings of words or phrases, meaning that is specialized to that context is conveyed. Teachers presented with an academic text tend to identify vocabulary that can obscure meaning as only multisyllabic words, instead of also including common multiple-meaning words, passive verb constructions, intricate processing terms, or connecting words that convey relationships within the list of vocabulary that might necessitate instruction (Francis et al., 2006; Kessler et al., 1985; Monaghan, 1999; Zwiers, 2008).

An example of vocabulary in the language of mathematics that might obscure meaning for ELs is when mathematical terms combine to form a new concept, increasing the task of comprehension. The phrase *a quarter of the apples*, takes the mathematical definition of *quarter* as referring to one-fourth of a whole and layers on complexity by needing to identify the fractional part of a set of whole numbers. Other examples are everyday words that also possess a unique mathematical meaning, such as *and* being used as a conjunction, but also used to indicate addition, *two and two are four*, or the combination of *least common multiple* (Irujo, 2007). Martiniello (2008) found that lack of familiarity with content and academic terms combined with a lack of proficiency in everyday vocabulary, or terms typically used in daily interactions, negatively affected ELs' performance on a mathematics assessment.

Content vocabulary names the terms and expressions specific to the academic discipline and is found in content area textbooks and other technical writing in the domain (Butler et al., 2004b; Joseph, 2012). Knowledge of content vocabulary is one element that allows students to engage with, produce, and talk about texts that are used in school, but some research shows it can be the vocabulary most difficult to acquire in the mathematics context (Joseph, 2012; Shaftel et al., 2006; Zwiers, 2008). The meaning of a content-specific term either needs to be explicitly taught or inferred from multiple supports; because content vocabulary is specific to that academic register, it may be context-reduced (Cummins, 1999; Joseph, 2012; Zwiers, 2008). Particularly in mathematics textbooks, a linguistic support might include text features that signify when a new term is introduced, such as bolding the term or including it in a label for a picture or diagram. Explicit instruction can also cause comprehension difficulty; when a teacher is focused on relaying the meaning of specific words in the mathematics classroom, the students may become confused as they struggle to correctly apply new forms of the content vocabulary (Deen & Hacquebord, 2002). In part, this problem may be caused by the students' tendency to apply their new learning to the context of their daily life, instead of a mathematical context. Understanding content vocabulary is complex yet essential to learning the new content or skills as the terms are used specifically in the domain.

Academic vocabulary refers to terms that appear across different content areas but vary in meaning based on their use in the different disciplines (Butler et al., 2004b; Joseph, 2012; Zwiers, 2008). Because their meaning depends on the context, the definitions can be difficult to acquire; academic language serves as the utility, or processing, terms that give context and purpose to content-specific terms (Cummins, 1999; Joseph, 2012; Zwiers, 2008). Defining the vocabulary terms that are used in a text to signal when a student needs to explain, solve, or discuss can be an important part of language objectives for instruction, as academic language allows students to describe complex ideas, higher-order thinking processes, and abstract concepts (Butler et al., 2004b; Francis et al., 2006; Kessler et al., 1985; Zwiers, 2008). Some research suggests that ELs transitioned into a mainstream classroom might encounter a teacher who does not promote academic language skills and therefore these students will take longer to catch up academically (Cummins, 1999; Francis et al., 2006). Academic language has received more attention of late, but its use, particularly with other types of vocabulary, still needs to be an instructional focus (Butler et al., 2004b; Francis et al., 2006; Zwiers, 2008).

Everyday vocabulary, as it is utilized in the language of mathematics, can also take on meanings specific to the mathematics context, such as *equal, rational, column,* and *table* (Dale & Cuevas, 1992; Zwiers, 2008). The use of everyday language in a mathematics register is not a lack of being mathematically precise, but is instead a means to communicate and make sense of mathematical meanings (Moschkovich, 2012). Barwell (2005) identifies ambiguity in the language used in the mathematics classroom, which is created when everyday vocabulary is used in conjunction with academic vocabulary in unfamiliar ways. This can be the problem with ‘formal’ or academic interactions, as everyday language is needed to convey and explain more explicit academic terms. Monaghan (1999) found in his analysis of a mathematics corpus that over half of the occurrences of *diagonal* presented the everyday, non-technical definition of *oblique,* instead of the technical, or content definition. Butler et al. (2004a) also found that 30% of the word tokens in their analysis of fifth grade social studies, science, and mathematics texts are frequently used words, instead of terms unique to the academic context. These studies serve as a note for mathematics teachers to be aware of the prevalence of everyday English, and the need to understand its different uses in a technical context.

Combined or overlapping meanings. As syntax influences meaning, definitions that come from combining terms or changing the context influence ELs' ability to acquire the vocabulary of the language of mathematics. When words are used consistently together, they are defined as lexical bundles, a group of three or more words that appear frequently in a specific register (Herbel-Eisenmann, Wagner, & Cortes, 2010). These bundles then take on meanings unique to the combination and the context used. Therefore, vocabulary terms need to be learned not only as singular words, but also in relationship with one another, as the meaning is unique to the context in which a particular mathematical expression is used (Dale & Cuevas, 1992). An example is the use of *right* as part of the angle name; the conventionalized form is *the right angle*, instead of the descriptive form, *the angle that is right*, which would not be considered the standard way to refer to an angle of 90 degrees (Schleppegrell, 2007). Teachers might not be aware of the specific linguistic patterning unique to certain mathematical terms, and therefore might also not provide explicit instruction or practice of the technical terms in the mathematical context of their use.

The interaction of academic and everyday terms becomes more complicated when both sets contain vocabulary with multiple meanings, leading to increased difficulty in acquisition and application for learners (Schleppegrell, 2007). *Polysemy* is used to describe words that have two or more different, but sometimes related, meanings (Lager, 2006). Specific language patterns may be associated with polysemous words, and therefore students need to learn not only the mathematical words but also how to apply them to convey meaning in mathematics (Schleppegrell, 2007). For example, again taking an angle measuring 90 degrees, most mathematicians use the term *right angle* to describe it. *Right* is a polysemous word because it can mean *correct* or it can describe an angle measuring 90 degrees. Since both mathematical and everyday languages are typically used in a word problem, it is up to the problem solver to distinguish when a word is being used mathematically and when it is not (Mestre, 2013). Other examples of polysemous words that can occur in such problems include: *volume*, *ruler*, *base*, *yard*, *face*, and *fair* (Adams, 2003). Because ELs are still learning various vocabulary words, they are likely to assign the more familiar meaning of a polysemous word, as is seen in the results of Martiniello's (2008) differential item functioning analysis of a fourth grade mathematics comprehensive assessment. The purpose of differential item functioning is to pull apart differing item difficulties for two groups of students with similar ability level. Reasoning that items with high linguistic demands might measure language competencies more than mathematical skills, this type of analysis determines which specific academic language features impact comprehension. Martiniello's (2008) research shows ELs with equivalent mathematical proficiency scoring lower than non-EL peers because of the linguistic load.

Homophones can also cause confusion, particularly when mathematics communication occurs orally. Take for example a word problem that ends with: *What is the sum of boys and girls?* Hearing *some*, and identifying it as part of a whole, instead of *sum*, and recognizing the instruction to find the added total, can lead to undue confusion.

Reduced speech can also cause confusion, as both *half* and *have* might sound like /haf/ in the phrase *one half* (Adams, 2003). All of these uses of multiple meaning words require not only awareness of the meanings, but also that students have instruction around the multiple meanings of words and how to identify combinations, decode the context, and apply the appropriate meaning (Butler et al., 2004b).

Symbols as vocabulary. Comprehension of the symbol system is another hurdle in understanding mathematical representation (Cocking & Chipman, 1988; Zwiers, 2008). Symbols in any language communicate meanings and messages, often unique to the language in which they are used. For example, in the mathematics register in other countries around the world, a comma is typically used to separate a whole number from a fractional part; in the English language of mathematics, a decimal point is used and a comma indicates separation of whole number place value (Dale & Cuevas, 1992). Geometric proofs are another area where symbols vary in meaning based on their use in different languages. The fact that these symbols can take on different meanings and values adds to the challenge of the language of mathematics (Dale & Cuevas, 1992; Zwiers, 2008). Understanding the particular meaning of symbols in the English mathematics register is important because symbols convey the relationship and variations among mathematical elements. The meaning of the symbols also expands when used with different operational processes (de Oliveira & Cheng, 2011).

Comprehending how symbols are used includes understanding their organization and management (Adams, 2003; Dale & Cuevas, 1992). For example, the reading of symbolic texts needs to be taken as a whole instead of in the traditional left-to-right (Adams, 2003). In the following: $(5 \times (4 + 2))$ the parentheses convey the order of operations, which are indicated by the $+$ and $\times$. While this is read as *5 times the sum of 4 and 2*, reading the text as words needs to be taught because of the lack of one-to-one correspondence between symbols and the terms they represent (Dale & Cuevas, 1992). Symbolic notation underlies the language of mathematics and possesses its own grammatical structure. Consider the numerical sentence: $2 < x < 8$ which is grammatical, but $2 < x > 8$ might or might not be and $2 > x > 8$ is definitely not. Students' proficiency with the symbolic aspect of the language of mathematics also influences their problem-solving performance (Mestre, 2013). Finally, comprehending the meaning produced by the interaction of words, numerals, and symbols is a skill in and of itself. *The number x is 4 less than the number y* could be incorrectly translated from left to right as $x = 4 - y$ instead of recognizing *less than* as $x = y - 4$ (Dale & Cuevas, 1992; Irujo, 2007). This added layer of convolution to the syntax of mathematics further demonstrates why explicit instruction of both syntax and vocabulary is needed to assist students in fully acquiring this language.

Written Discourse

Another intricate feature of the language of mathematics is its written discourse, comprised of register-specific text organization, visual content and layout, extended sentence length, and the unique genre of word problems. The written command *explain* beneath an algebraic expression takes on a very different meaning than the setting of a parent pointing to a broken vase. The mathematics principles and concepts conveyed through text should be enhanced by instruction in order to support the mathematical understanding and ownership of learners (Barwell, 2005). Understanding the language of mathematics involves the written pattern of discourse; the meaning of a sentence, let alone a term, is influenced by the surrounding text features.

Text organization. Reading a mathematics textbook is taxing not only because of the linguistic aspects of vocabulary, syntax, multiple meanings of words, and mathematical symbols, but also because of spatial positioning of numbers, symbols, and text that combine content and context. The pages are conceptually packed and arranged up-and-down and left-to-right, although the reading of symbolic “sentences” is not always strictly left to right, but involves knowledge of possible combinations (Lager, 2006). Mathematical texts are often presented in a procedural format; students often view all mathematics as procedural in nature and therefore overlook other aspects, such as instruction and explanation (de Oliveira & Cheng, 2011). Other linguistic aspects of textbooks, such as the use of imperatives, can convey the text itself as authoritative (Butler et al., 2004b); roles and relationships are therefore conveyed through the text (de Oliveira & Cheng, 2011). The process of reading a mathematics text, even without the influence of discourse around the text, combines previously learned mathematics concepts, procedures, and applications, and the knowledge of which to apply. Relevant background knowledge also needs to be accessed and applied along with mathematical thinking processes. All of these factors lead to the necessity of a slower reading rate and multiple readings in order to fully understand the mathematical content (Dale & Cuevas, 1992).

The use of illustrations and visuals provides further support to verbal statements by often representing the context in a more explicit way. The visual elements of mathematics often represent dynamic, multiple time frames which employ spatial knowledge as well as understanding of representation of real-world objects or mathematical content (de Oliveira & Cheng, 2011). Visual elements add another layer of legibility to the text, but students have to understand how to use and interpret them. The use of a visual becomes more complicated as the mathematical information often foregrounds and frames one aspect, while irrelevant information is reduced or eliminated. Take for example a visual of a puzzle and a question asking: *What fraction of the puzzle is not completed?* Students would need to ignore background information on the concepts of puzzles and instead frame the example around the perspective of fractions of a whole. Because of this aspect of the textual features in a mathematics textbook, the intent, interests, and goals of students may impact their interpretation of the signs and language presented (de Oliveira & Cheng, 2011).

Finally, the layout of the text itself might present confusion, as the syntactic boundaries of clauses are often not indicated by the printed text. For example, Martiniello (2008) found in her differential item functioning analysis of a fourth grade mathematics comprehensive assessment that learners often interpreted complex sentences based on their print layout, such as associating the line *number on a spinner identical to the one* with the numeral one on a spinner instead of comprehending the overall structure: *To win a game, Tamika must spin an even number on a spinner identical to the one shown below* (Martiniello, 2008, p. 342). Her analysis fit with other research she studied on the visual-syntactic text formatting and its impact on reading comprehension (Martiniello, 2008). Overall, the application of unusual structures such as diagrams, visuals, and the arrangement of the text itself can all lead to unique meaning in the discourse of mathematics.

Sentence length adds complexity. Not only the layout of the text, but also the length of a sentence affects comprehension of text present in the mathematics register (Zwiers, 2008). Most research on text comprehensibility has found a direct correlation between sentence length and overall linguistic complexity, resulting in difficulty of comprehension of longer sentences (Abedi & Lord, 2001; Butler et al., 2004b; Martiniello, 2008). Bergqvist et al. (2012) identified in their analysis of a corpus of vocabulary in mathematical tasks that the total number of words per each sentence in a task is directly related to the task difficulty. Abedi and Lord (2001) found in their research of an 8^{th} grade mathematics assessment that the length of a text is one of many linguistic factors that can inhibit ELs from performing to their actual level of mathematical ability. As students tend to have a global reading strategy in which they skip over words that they do not know (Deen & Hacquebord, 2002), this can also affect their comprehension of word problems. Finally, the attempt to decode one word can affect comprehension of the whole; the more time and effort a reader puts into decoding one word, the less likely they are to remember the preceding words of the phrase and therefore be able to combine and comprehend multiple meanings (Martiniello, 2008).

Word problems. One specific example of the contextualized meaning found in the written language of mathematics is the common word problem. Defined as a situational problem that uses multiple phrases to lead the solver through one or many steps to its solution, word problems are also called story problems, although their story is separate from time and space. In themselves, mathematical word problems are focused examples of the use of mathematics discourse, in that they combine the more specialized aspects of the language while suspending a typical narrative focus on details and instead direct everything towards a mathematical solution (Butler et al., 2004b; Dale & Cuevas, 1992). Sweeney (2014) explored the linguistic complexity of comparison situations, such as: “Kyle ran 6 miles this morning. This is two more miles than Amy ran. How many miles did Amy run?” (p. 25). She found in her analysis that complexity increased from 3rd to 5th grade. As the complexity increases, so does the need for instruction around solving problems presented in this particular structure.

Most word problems follow a three-part compositional structure: 1) a “*set-up*” establishing characters and location, 2) *information* giving what is needed to solve the problem, and 3) a *question*. These three parts can be combined into one sentence by using subordinate clauses or subjective structures, such as *If…then…*. This composition of word problems is more similar to algorithms or algebraic problems than the conventions of oral or written storytelling. In word problems, information is given to convey relationship among the parts that are often conveyed through variables and symbols, a feature similar to algebraic problems (Fang & Schleppegrell, 2010; Gerofsky, 1996; Sweeney, 2014).

Word problems also convey multiple layers of meaning, having locutionary (literal meaning), illocutionary (performative intention), and perlocutionary (effect upon the audience) affects. Because the referents in word problems seldom exist in real life, the deixis, or indication of words, can cause locutionary problems. The illocutionary intention of solving the problem assumes that it is solvable, that no other information is needed apart from mathematical knowledge the student already has access to, and that there is one right answer which the teacher can determine as correct and which can be represented in mathematical or algebraic (numerical) form (Gerofsky, 1996). From an outside perspective, word problems can be analyzed for what they say about the world (the experiential meaning), in regards to the social relationship they convey (the interpersonal meaning) and the way they weave meaning into a message (the textual meaning) (Fang & Schleppegrell, 2010).

All of these assumptions around the specific written discourse of word problems can be hidden from students and add yet another layer of confusion. Gerofsky (1996), in her linguistic analysis of word problems, summed up this tension by describing how word problems are hypothetical and do not reflect real-life situations because there is no extraneous information or authentic situations that require problem solving. Providing instruction around syntax and vocabulary could help the comprehension of word problems, but the overall unique discourse pattern continues to present problems of its own.

English Learners' Access to the Language of Mathematics

In school, language is the vehicle of learning and instruction (Dale & Cuevas, 1992). For ELs, the challenge of learning mathematics is compounded by simultaneously having to acquire English. Research shows a strong correlation between reading skills and mathematics achievement, particularly around more language-centered tasks, and recent evidence shows the correlation might be even stronger for ELs (Abedi & Lord, 2001; Bergqvist et al., 2012; Dale & Cuevas, 1992; Francis et al., 2006; Martiniello, 2008). In order to learn mathematics through a second language, learners must first reach a minimal level of proficiency in the cognitive academic skills needed to understand mathematics as well as the language skills needed to convey the mathematical content (Dale & Cuevas, 1992). Learners can quickly become frustrated if new mathematics content is not initially accessible or understandable (Lager, 2006). With federal accountability mandates of the previous No Child Left Behind Act, mathematics instructors noted data disaggregated for ELs that identifies this group showing a lack of progress and scoring significantly behind non-EL peers (NAEP, 2013). Because of differing linguistic proficiency, students have varying access to the modes of communication in the mathematics classroom, which affects how they are able to demonstrate and build their knowledge of mathematics.

The questions in texts and tests, classroom discourse, and what is perceived as legitimate knowledge are all common aspects of communication that affect a student's access to mathematics (Zevenbergen, 2000). The United Nations Educational, Scientific, and Cultural Organization stated in their 1975 report on mathematics education that the main area of difficulty in the mathematics classroom is that the teacher already understands and takes for granted the language of mathematics; instead of providing comprehensive language instruction, typically only the content vocabulary is taught. This organization has recommended that ESL teachers and mathematics educators collaborate to present instruction that views the language of mathematics as a subset of the mathematics content that needs to be taught (Monaghan, 1999). The moves a teacher makes can not only support English language learning in the mathematics register, but also help students learn how to access previous knowledge of language and content.

Impact of Linguistic Complexity

As previously demonstrated, the language of mathematics has complex linguistic features, which are compounded for English language learners. The mix of academic, everyday, and content vocabulary all specific to a mathematical context increase the difficulty of its comprehension and acquisition. Because complex linguistic forms in mathematics textbooks can compound difficulty in comprehension, students come to view the textbook as an additional learning obstacle instead of a learning tool (Lager, 2006). Shaftel et al. (2006) have analyzed the language characteristics in mathematics test items and their specific impact on the performance of ELs at 4th, 7th, and 10th grade. Their differential item analysis found that at grade 4, prepositions, ambiguity, polysemous words, complex verbs, pronouns, and math vocabulary presented the most difficulty for all elementary students, not just ELs. As the grade level increased, the linguistic elements presented fewer difficulties, although mathematics vocabulary remained consistently difficult. This echoes findings that the intricate language in word problems has a greater impact on lower level learners (Shaftel et al., 2006). While not all language learners are at lower levels mathematically, their ability to demonstrate their actual knowledge is often hindered because of the impact of language.

Haag et al. (2013) analyzed data from a state-wide mathematics assessment in Berlin to determine the interrelationship between academic language features such as text length, general academic vocabulary, and the number of noun phrases and the performance of language learners compared to native speakers as measured by differential item functioning. Their analysis demonstrates that the more academic language features an item contains, the more ELs have difficulty in understanding it, with grammatical and lexical features having a higher impact on comprehensibility than specific descriptive features. Their conclusion fits with classroom practice, as mathematical content terms are often explicitly taught whereas other academic language features are instructed less frequently and systematically. They concluded that there are several academic language features that contribute to overall performance of ELs, with noun phrases being a more prominent factor than the others (Haag et al., 2013).

Academic language is difficult to define and researchers vary on what should be learned by ELs in order for them to be proficient in academic vocabulary (Browne, 2014; Butler et al., 2004a; Butler et al., 2004b; Coxhead, 2000; Zwiers, 2008). Coxhead (2000) created an *Academic Word List in the Academic Corpus* by identifying the most frequent word families for repeated words in the corpus across content areas. This list of 570 word families gives academic terms that are not subject-specific vocabulary and removes everyday words as identified through the General Service List. The goal of this list of academic vocabulary is to guide instruction to support ELs in reading and writing, as learning these terms would allow for comprehension of about 10% of all vocabulary in academic text (Coxhead, 2000). Other studies conducted by Butler et al. (2004a, 2004b) attempt to identify what academic language is used in upper elementary textbooks and assessments in order to provide instructional recommendations on academic vocabulary and better develop content and language assessments for ELs. The first study categorized vocabulary in the textbooks, searching for specialized (within a content area) and general (across content areas) academic terms. The researchers identified only 15 general academic words between fifth grade social studies, science, and mathematics textbooks (Butler et al., 2004a). The second study compared the language used in texts to the language in content standards, but did not find a correlation between the language used and academic language (Butler et al., 2004b).

Browne, Culligan, and Phillips (Browne, 2014) took a different approach and looked to identify everyday words by analyzing a corpus of 273 million words composed of various forms of oral and written discourse, such as journals, television, fictional texts, procedural documents, and formal speeches. They created the New General Service List (NGSL) from the most frequent words across this variety of text and oral discourse, building off an initial research model that created the General Service List from a corpus of around 2.5 million words. The goal of the NGSL is to help language learners and teachers know which words should be learned to more quickly acquire English, by identifying the most important high-frequency words that give the greatest coverage of English texts with the fewest possible words (Browne, 2014). With research around lists of everyday and academic vocabulary, and content vocabulary often identified in the glossary of a textbook (Fong et al., 2009), no one has yet investigated how prevalent the different types of vocabulary are in the directional sentences in an upper elementary mathematics textbook.

Providing Access

Given the intricacy of the language of mathematics, some researchers argue the benefit of assessing ELs mathematical skills without the hindrance of language. Abedi and Lord (2001) explored the impact of reducing linguistic complexity on items in a national mathematics assessment; they modified test items to lower linguistic complexity and then compared ELs’ performance to non-language learning peers on both assessment versions. The average scores of ELs and students from lower socioeconomic status were lower than the mean of the majority of students, but on the linguistically modified test items their mean scores were slightly higher. Lager (2006), who also explores the connection between language complexity and performance on mathematics assessments, found that ELs did not perform as well as non-EL peers. When interviewed, the students identified various linguistic aspects of the problems such as academic vocabulary, complex noun phrases, polysemous terms, and the use of mathematical symbols that led to confusion and lack of performance (Lager, 2006). Removing these features to allow for increased student access, however, will eventually hinder ELs' mathematical development; as research shows, the language of mathematics is thoroughly integrated into mathematical content knowledge and skills, and ELs ultimately still need to acquire it (Abedi & Lord, 2001; Lager, 2006; Martiniello, 2008).

English learners need to have access to learning in ways that promote academic language development in the language of instruction. The development of academic language should occur in the context of learning the discipline; subject-area teachers need knowledge of how language works in their discipline as well as the metalanguage needed to make that knowledge accessible to students (Achugar et al., 2007). The tension of a multilingual classroom to promote mathematical understanding has been labeled the 'teaching dilemma,' where teachers want to teach explicitly yet allow exploration of the intricate and dialectical aspects of mathematics (Adler, 1998). This instructional tension has been further broken down to include the dilemma of code-switching (when to allow use of home language to build new language), the dilemma of mediation (when to provide correction), and the dilemma of transparency (when to highlight specific language features) (Adler, 1998, p. 26). While instructors might tend to over- or underemphasize one aspect more than others, overall there is a need to value the first language, provide feedback that promotes language development, and instruct learners on language use.

One way of providing ELs access to content is to increase reading and comprehension skills first. Although students might attempt to apply comprehension strategies commonly used in other academic subjects, such as figuring out unknown terms from the context, word problems in math are typically short, lacking context and the natural redundancy or repetition of language. Strategies around clue words or selection of an operation based on the size of the numbers can also prove faulty, as the multiple meanings of mathematical terms and the variation between words taken singularly or combined into terms can vary depending on use (Irujo, 2007). The solution, according to Francis et al. (2006) is explicit instruction, as academic language is difficult for ELs to acquire without explicit integration into content curriculum. The aspects of linguistic complexity, as well as specific needs of ELs, combine to create a situation where explicit language instruction may be needed to provide access to content.

The Gap

The research reviewed above demonstrates the many aspects of the language of mathematics that present difficulty for ELs. The syntax, semantics, and discourse factors found in this language all affect ELs' mastery of mathematics in their second language. For mastery to happen, instruction needs to occur. Most of the research studies specific to the impact of language on ELs' acquisition of mathematics have examined the language features present in word problems in comprehensive assessments and textbooks (Abedi & Lord, 2001; Bergqvist et al., 2012; Butler et al., 2004a; Butler et al., 2004b; Deen & Hacquebord, 2002; Gerofsky, 1996; Haag et al., 2013; Lager, 2006; Martiniello, 2008; de Oliveira & Cheng, 2011; Shaftel et al., 2006; Sweeney, 2014). The written explanations, directions, and questions used in a mathematics text, however, remain unexplored.

Looking specifically at directions in mathematics textbooks, these also contain difficult linguistic features and vocabulary of the language of mathematics. Students are most likely to interact with this section independently, as they apply the content learned from the instructor to the practice tasks and questions outlined in this area of the text. These directions in a mathematics textbook, while sometimes used to guide the solution of problems, are typically without the context of a word problem, or even the length typical in the directions of a word problem; this important area lacks research. Research has shown that the length of a word problem correlates with the difficulty a student has in solving it (Abedi & Lord, 2001; Bergqvist et al., 2012; Butler et al., 2004a; Butler et al., 2004b; Martiniello, 2008), and that the way word problems are written is not realistic or representative of real-world problems (Fang & Schleppegrell, 2010; Gerofsky, 1996). Similar analysis has not been conducted on language used for giving directions, whether the instruction is an actual question or an indirect command. Since none of the studies on the linguistic features of mathematical texts mentioned above have analyzed this language in a mathematics textbook, it should prove insightful to look at the linguistic features and vocabulary used in instructions, questions, and commands in order to help ELs better acquire the language of mathematics.

Research Questions

This review of the research, as well as the personal experiences described in the first chapter, lead to the following questions. What grammatical and syntactical features of the language of mathematics that researchers identify as difficult to understand are commonly used in conveying directions in an upper elementary mathematics textbook? How do the words and lexical bundles that are used for directions compare to the vocabulary recommended for instruction? The process of answering these questions starts with analyzing the linguistic features in an upper elementary mathematics textbook and comparing them to aspects of the language of mathematics that researchers identify as more difficult to acquire; ultimately, this should indicate features that can be explicitly taught to ELs in order to help them achieve proficiency in the language of mathematics.

Summary

In this section, research on the specific features in the language of mathematics was reviewed, specifically the aspects of the-syntax, semantics, and written discourse features that present difficulties for ELs. The needs of ELs in learning mathematical content was addressed, as well as the impact of having to learn academic and content vocabulary. As the need for research in language development of the language of mathematics has been identified, specifically analyzing the language aspects of directions in a mathematics curriculum, the next chapter will discuss the methodology for this study.

CHAPTER THREE: METHODOLOGY

This study takes the linguistic characteristics identified by researchers as challenging in the language of mathematics, and uses them to identify features of the language present in the directional sentences in an upper elementary mathematics textbook. A process was identified that selected and labeled that text, input it into a corpus, and then identified some of the grammatical and syntactical linguistic features previously named by researchers. This process led to evidence that can be used to answer the research questions: What grammatical and syntactical features of the language of mathematics that researchers identify as difficult to understand are commonly used in conveying directions in an upper elementary mathematics textbook? How do the words and lexical bundles that are used for directions and the recommended vocabulary for instruction compare? The overall goal was to identify features present in written directions that should be explicitly taught to ELs in order to help them achieve proficiency in the language of mathematics.

This chapter begins with an overview of the quantitative research paradigm. Then a more detailed description of the data collection will be outlined, including how the text was selected, which linguistic features identified by researchers as adding complexity to the language of mathematics were labeled in the corpus, which programs were used for this categorization. The pilot study will be described, along with the resulting adjustments to the data collection. A plan for ensuring validity of this research study will also be described.

Quantitative Text Analysis Research Paradigm

Quantitative research strives to determine a relationship between or within variables. Specifically, it looks at co-occurrence across data sets (Biber & Conrad, 2001; Mackey & Gass, 2005). Analyzing the data in a linguistic corpus allows for an investigation of language patterns. Teachers might be unsure which linguistic structures and vocabulary are important to teach; a corpus study removes the personal perspective and offers quantitative linguistic data, which can then be analyzed for connections between grammatical features and items specific to the register (Biber & Conrad, 2001).

The research questions for this study focus on identifying linguistic features in the directions of mathematics that might prove difficult for ELs to acquire, based on relevant research. By creating a corpus from the sentences that give directions in the fourth grade *Math in Focus* textbook, this provided a means to identify prevalent lexical bundles and patterns of linguistic features common to the language of mathematics. The quantitative data categorized from the corpus was then analyzed to identify patterns that might be present in a larger study, therefore informing instruction (Biber & Conrad, 2001).

Data Collection

Text Selection

As this study analyzes linguistic features, a curriculum that has been identified as language-heavy is essential (Fong, 2009). The *Math in Focus* curriculum was identified as one that focuses on word problems; as a result, the entire text is often challenging and dense because of the syntax and vocabulary used. While previous research has been conducted on the linguistic complexity of word problems, very few research studies have been built around the other text in mathematics textbooks (Bergqvist et al., 2012; Joseph, 2012; Monaghan, 1999), and none that explicitly analyze the linguistic features present within a mathematics textbook. Because students interact independently most often with a mathematics textbook when they are trying to solve problems or complete a task, a logical next step for analysis is to analyze the instructions and directions that are given for completing various exercises in the text. These sections require students to process the information presented, as well as carry out the activity or task within, without the influence of the teacher mediating their understanding (Butler et al., 2004b). Therefore, the text selected for the corpus in the study is composed of all sentences in the fourth grade student textbook that give directions that require student action. The fourth grade level text was chosen in order to provide insight into the connection between elementary and middle school demand in the level of mathematics, as well as to acknowledge the shift that starts in third grade to focus on more academic content in textbooks rather than the development of reading skills (Butler et al., 2004b).

To compile the corpus, the sentences in the fourth grade *Math in Focus* textbook sections that direct students to answer a question or follow a command were transcribed into a corpus and labeled with the specific location and type of text section. The *Math in Focus* textbook is structured so that fourteen different topics are presented in separate chapters, split between Book A and Book B so that the size of the text remains manageable for student use. Each chapter is focused on a specific mathematical skill and aligned, to some extent, with the domains in the Mathematics Common Core State

Standards. Since the curriculum was originally adapted from Singapore and distributed in the United States before the wide-spread implementation of the Common Core State Standards, some of the chapter topics are no longer covered in that specific grade level under the Common Core and therefore most districts following the CCSS view them as supplemental (Fong, 2009). Because of this, only the chapters that are aligned with that fourth grade level CCSS are included in this study, with each transcribed entry also tagged to the Common Core domain as identified by the lesson standard focus (Appendix B). For example, the first lesson in *Math in Focus* Grade 4 Book A is titled *Numbers to 100,000*. This lesson is within the chapter on *Place Value of Whole Numbers*, and covers the standards under the Common Core domain of *Number & Operations in Base Ten*. While some of the instructions might be for a task that requires algebraic thinking, a separate domain in the Common Core, because the lesson was identified under the theme of number sense, the directions therein contain the same label for ease of categorization. This also allows any patterns that might exist around one skill or mathematical domain and not another to emerge in the corpus, such as the previously identified linguistic complexity in the data analysis, statistics, and probability domain (Martiniello, 2008) or the variation of linguistic features between domains (Butler et al., 2004a).

Within each *Math in Focus* chapter, the first lesson is an introduction that reviews background knowledge. This is followed by the content lessons; each focuses on one concept or skill and includes examples and labeling of new terms (in a *Learn* section) before presenting guided and independent practice problems. Each chapter then ends with a section for review of all the lessons and test preparation for the chapter assessment. Throughout the chapter, additional activities, games, and challenge problems are included in the various lessons to allow students to practice the concepts through hands-on activities or ones that require higher-order thinking skills. The headings for all of the types of instructional text found in the chapter, composed of sections where students need to follow directions, were labeled with heading initials (Appendix C). Each sentence entered into the corpus was labeled with the heading initials as well as with another label indicating the specific location in the textbook (e.g. *4A.1.1.p9* refers to Grade 4 Book A Chapter 1 Lesson 1 page 9). This is not only so that each sentence can be located again if needed, but also so that any linguistic patterns identified can subsequently be analyzed according to the type of activity indicated by the heading initials or location in the domain within the curriculum. Any problems in the *Guided Practice* or *Independent Practice* that embody the genre of word problems were not included. Because word problems contain unique linguistic features within their sub-genre of the language of mathematics, removing them allows for a clearer identification of the language used for directions and not that used to create a situation within its own context. All of these decisions were undertaken to allow patterns to emerge as well as to ultimately identify linguistic features unique to the mathematics domain which might benefit ELs if taught through explicit instruction.

Categorization of Text

The list for text categorization of the corpus entered into Microsoft Excel was created based on the previous research and linguistic analysis studies, and then adjusted according to results from the pilot study. Microsoft Excel allows for calculations of frequency and average of occurrence within the corpus and percentage of representation out of all directional sentences or phrases. These calculations were conducted on the entire corpus and then on the five separate CCSS domains. In order to complete these calculations, each sentence giving direction within the fourth grade *Math in Focus* textbook was labeled in Microsoft Excel according to the following list of linguistic features:

1. Type of sentence (question, command, or statement)
2. Verb-initial sentence (yes/no)
3. Verb tense
4. Passive verb form (yes/no)
5. Number of verb phrases
6. Length of verb phrase(s)
7. Number of nominal phrases
8. Length of nominal phrase(s)
9. Length of sentence
10. Number of relative clauses

While these categories were identified before the entire corpus was recorded or analyzed, they were designed so that the data on occurrence could be compared to the results of previous research studies and allow for the emergence of other patterns. See Appendix E for specific examples of each linguistic features that is described below.

Type of sentence. Sentences are described as a linguistic unit that is composed of a single independent clause or a number of clauses joined together. The tenor of the sentence determines the interaction; sentences are commonly categorized according to their pattern for interaction as a statement, question, command, or offer (Derewianka, 2013). Identifying the form and function of this mode of communication can better help teachers and students understand the dynamics of language and how it is influenced by its context. The sentences for instruction in *Math in Focus* were categorized as statement (*S*), question (*Q*), or command (*C*). This categorization came as a result of the pilot study, as the majority of sentences transcribed from the first chapter of the fourth grade Book A are in the form of a command. Some of the commands are conditioned by modals or directed to a third party, such as: *You can use place-value charts to help you compare these decimals* and *Player 1 rolls the number cube two times to get two numbers* (Fong et al., 2009, 4B p. 29, 4A p. 246); these were also included under the label of command (*C*) as they make statements that imply what a person should do. Further analysis will show whether the pattern of frequent commands continues throughout the text and across grade levels.

Verb-initial sentence. The appearance of verb-initial sentences appears to be related to the prevalence of commands in the instructional text. More commonly used in oral interaction than in written text, verb-initial sentences are a form used to give instructions or directions (Derewianka, 2013). When a sentence begins with a verb, the subject most often implied is *you*, and can therefore be more difficult to understand, particularly if language learners are relying on a common subject–verb–object pattern. Because of its frequency of occurrence in the pilot, verb-initiality was also recorded.

Verb form, phrase frequency, and length. Verbs convey the action in a sentence, as well as indicating processes around doing, thinking, saying, relating, and existing. Conveying tense and aspects of time, English contains three tenses (present, past, and future), and four aspects (simple, perfect, progressive, and perfect progressive). With that, some verbs can be active or passive. Research shows that passive verbs and auxiliary terms typically cause more difficulty for understanding (Abedi & Lord, 2001; Butler et al., 2004a; Fang & Schleppegrell, 2010; Irujo, 2007). Depending on the tense and aspect, a verb group can include one term or many, along with auxiliary terms, modals, negatives, prepositions, or adverbs (Derewianka, 2013). Sentences can be compound, complex, or compound complex, which would mean more than one verb phrase would be present in a sentence. The identification of verb tense was recorded, focusing on the main verb in the sentence. Since more auxiliary terms can increase difficulty of comprehension, the count of all the words in a verb phrase and the total number of verb phrases in a sentence was also recorded.

Frequency of nominal phrases and length. As multiple words can form a verb group, multiple words can also be part of a noun, or nominal, group. The head noun is typically a person or thing (represented by a noun or a pronoun) and can have pre-modifiers and post-modifiers. Modifiers can include articles, demonstratives, possessives, or adjectives in the form of quantifiers, describers or classifiers. These terms combine to give more information regarding the head noun or pronoun. Nominalization can also transform verb forms, adjectives, and adverbs into nouns as it presents actions, processes, states, and circumstances (Butler et al., 2004a). Nominal groups should be viewed as a chunk of information rather than as individual words (Derewianka, 2013). Academic texts often use nominalization and complex noun phrases to condense text, requiring the reader to understand and incorporate it into background knowledge in order to use it as the subject or link one concept to another (Zwiers, 2008). This need for a long, complex nominal to be interpreted as a whole often proves difficult for ELs (Abedi & Lord, 2001; Butler et al., 2004a; Haag et al., 2013; de Oliveira & Cheng, 2011; Shaftel et al., 2006;

Sweeney, 2014). It was beneficial to record the length of various nominal groups present in directional sentences as an indication of comprehension difficulty. This length was then compared to the length of the sentence itself, to identify what percentage of a typical sentence is composed of one or more noun phrases. Finally, the length of sentences was analyzed itself, as research also indicates that the longer the sentence, the greater the linguistic complexity (Abedi & Lord, 2001; Bergqvist et al., 2012; Butler et al., 2004b; Martiniello, 2008).

Frequency of relative clauses. Relative clauses are a part of the noun group that qualifies the head noun. If the information they provide is essential to determine which things are referred to, they are called defining relative clauses; otherwise, they are non-defining relative clauses that add extra information (Derewianka, 2013). Relative clauses can prove difficult for ELs as understanding what is being described can be obscured in the composition of the clause (Abedi & Lord, 2001). Because of this, further analyzing the nominal group for the existence of relative clause could indicate an area for language instruction.

Comparison with Vocabulary Lists

Another purpose for creating the corpus was to compare the terms used in the directional sentences in the fourth grade *Math in Focus* textbook to the content and academic vocabulary recommended for mathematics instruction; this is to give an indication if ELs are likely to receive instruction on the vocabulary used in the textbook. Every state has content standards identified for mathematics instruction. With the movement of several states to adopt the Common Core State Standards, teachers have to use the provided curriculum to reach the determined standards. While there are not specific grade-level vocabulary lists identified, the Common Core provides a glossary of terms. This glossary explains the less familiar content vocabulary that is used to list the specific standards and give examples of application (*CCSS of Mathematics*, 2010). Because these are the standards for the state where this research study was conducted, as well as for the majority of the United States at this time, it is appropriate to include these vocabulary terms in order to compare and highlight differences between Common Core recommended content vocabulary and terms that are actually used.

Another source for content vocabulary terms to compare to the terms used in the directions analyzed is found in the curriculum itself. *Math in Focus* was not originally designed to meet the Common Core specifically, but with the adoption of the standards, the curriculum has identified the connection between the existing curriculum focus and the topics identified by the standards. Elementary curriculum typically also contains recommendations for vocabulary instruction. *Math in Focus* is no exception; these identified content vocabulary terms contain more textual features and support for instruction regarding acquisition, so comparing the terms recommended for teaching with those that are used in the text should provide a beneficial indicator for language instruction.

The language of mathematics contains more than content vocabulary. Academic terms are another layer in the language of mathematics. Zwiers (2008), promoting the teaching and use of academic language in the classroom, gives a Partial Academic Word List, which is a subsection of Coxhead's *Academic Word List in the Academic Corpus* (2000). These academic terms are found across various academic disciplines; by compiling this list, Zwiers intends to heighten awareness of academic terms that are used to convey abstract and complex concepts. Comparing this list of academic terms to the corpus of terms from the *Math in Focus* directions should provide more insight into the importance of explicit instruction around academic language. Because previous research demonstrates not only the need for a focus on academic language, but also the interaction of academic, everyday, and content vocabulary, it is important to analyze the occurrence of academic vocabulary in the directions in the fourth grade *Math in Focus* textbook (Achugar et al., 2007; Adams, 2003; Barwell, 2005; Butler et al., 2004b; Cummins, 1999; Dale & Cuevas, 1992; Francis et al., 2006; Herbel-Eisenmann et al., 2010; Irujo, 2007; Joseph, 2012; Kessler et al., 1985; Lager, 2006; Martiniello, 2008; Monaghan, 1999; de Oliveira & Cheng, 2011; Schleppegrell, 2007; Shaftel et al., 2006).

Data Compilation Instruments

With the large amount of text to be recorded, categorized, and analyzed, choosing the appropriate program could prove essential to the success of the study. Denzin and Lincoln (2000) have identified the benefit of using various computer programs to allow for the categorization and analysis of data. Each program in this study was selected based on ease of use as well as the ability to categorize, identify, and analyze the more prevalent linguistic features identified by researchers as problematic.

Microsoft Excel. The first program utilized in this study to analyze the corpus was Microsoft Excel. Each directional sentence was first typed into Microsoft Word, with the labeling of the specific location in the textbook as well as the heading for that section tab delineated; each entry was separated by a carriage return so that the file could be uploaded as a plain text file into Microsoft Excel and produce three columns: location, section heading, and text. Microsoft Excel is an appropriate platform for initial collection as it provides flexibility in categorization, ease of search, and ability to write functions that allow for identification of terms as well as resulting percentages. Values and labels were entered into subsequent columns indicating the presence or number of specific linguistic features; these specific features of data collection are described in the pilot study.

AntConc. The concatenated sentences were then run through AntConc (Anthony, 2015), a software that identifies the most frequent terms and lexical bundles present in a corpus. AntConc has been specifically created in order to support the concordancing and analyzing of linguistic data, including the feature of comparing the existing terms in the text entered to other word lists (Anthony, 2015). With this additional feature, the corpus of directional sentences can be compared to the CCSS recommended vocabulary, the vocabulary lists within the fourth grade *Math in Focus* curriculum, and Zwiers' (2008) Partial Academic Word List in order to identify which terms are recommended for instruction and which are missing. AntConc allows for various layers of reports to be run. After uploading the corpus from Excel, AntConc automatically identifies "clusters" or lexical bundles; the cluster size was set between three and five to fit with previous research reviewed (Herbel-Eisenmann et al., 2010). A section on "collocates" lists the most frequently used words, while selecting the "word list" feature allows the user to compare the corpus to any other defined word list or corpus. The word lists identified through the pilot study were then uploaded and compared to the words present in the corpus of directional sentences from *Math in Focus*.

Procedure

Once the text was identified as well as the means to collect and then analyze it, the next step was to determine which specific linguistic features to initially classify in Microsoft Excel. Since the purpose behind collecting and recording the directional sentences was to identify prevalent linguistic features, choosing which linguistic features to label when transcribing the text took careful consideration. Using a combination of the previously cited research and linguistic analysis studies (Abedi & Lord, 2001; Butler et al., 2004a; Butler et al., 2004b; Gerofsky, 1996; Haag et al., 2013; de Oliveira & Cheng, 2011; Shaftel et al., 2006; Sweeney, 2014), a pilot study was conducted, testing out the software as well as allowing for any initial patterns to be identified that would help categorize the corpus.

Pilot Study

The pilot study was conducted in order to try out the initial data collection, and identify the initial categories for linguistic features. First the sentences and location tags (both section and chapter, lesson, and page number) were recorded in Microsoft Word with tab- and carriage-delineated entry in order to allow for importation into two columns in Microsoft Excel. Only the sentences used to convey directions in the first chapter of *Math in Focus* Grade Four Book A were entered to allow for a larger focus on the process without the burden of extensive content. While the directions might be several sentences in length, each sentence was entered in a separate line to allow for individual analysis of the sentence as a whole (Appendix D). Once in Excel, the sentences were categorized in another column as a question (*Q*), command (*C*), or statement (*S*) based on their syntax. In order to have the ability to analyze various directional sentences under the different CCSS domains, each sentence was also labeled with the CCSS domain corresponding to the main focus of the lesson. Because of this, all the sentences in the lessons in Chapter 1 were labeled with *Numbers & Operations in Base Ten* (*NBT*) in another column in Excel. The presence of verb initiation (Yes = 1, No = 0) as well as the verb form were entered into two subsequent columns. In the pilot study it became apparent that multiple sentences had a verb as the first word; this verb initiation was indirectly identified as a complex linguistic structure in that referents can be hidden based on the implied subject (Gerofsky, 1996).

The next linguistic categories that were entered into Excel columns differ from the final ones. Initially the number of words in a noun phrase, the number of noun phrases, as well as the number of prepositions, modifiers, and articles were identified. But upon initial categorization, it was apparent that the number of all the individual types of terms would equal the number of terms in the noun phrase. While beneficial for verification of accuracy in labeling, the information seemed redundant and not readily apparent how a classroom teacher would benefit from knowing these various amounts. The previous research was revisited, and it was found that the frequent use of nominal phrases was mentioned as a difficult linguistic feature (Abedi & Lord, 2001; Butler et al., 2004a; Butler et al., 2004b; Haag et al., 2013; de Oliveira & Cheng, 2011; Shaftel et al., 2006; Sweeney, 2014). As a result, the above categories were abandoned in favor of generally identifying the number of nominal phrases and the length of each in Excel columns, as that research also shows a correlation between length and difficulty of language acquisition. Initially the number of articles and sequencing terms were also identified and again disregarded in favor of the identification of the presence and type of relative clauses whose difficulty is supported by research (Abedi & Lord, 2001; Butler et al., 2004a).

The pilot also provided a chance to explore the technology used, as various functions in Excel needed to be created and the AntConc (Anthony, 2015) program needed to be field-tested. The initial results from these programs provide data to support beginning observations. For example, the text analyzed for the pilot shows within the labeling in Excel that 82% of the directional sentences are commands, with 11% statements and 5% questions. 74% of those sentences also begin with a verb (per the percentage presence of Yes = 1 out of the whole), which means the subject is implied and fits the observed form of commands. Analyzing the corpus as a whole in AntConc, the five most frequent words are: *the*, *number*(*s*), *each*, *in*, and *form*. The five most frequent lexical bundles are: *look at the numbers*, *express each number in*, *order the numbers from*, *the value of*, and *at the numbers in the*, and subsequent variations or smaller bundles within those listed. Some of these terms seemed to be specific to the domain, *Numbers & Operations in Base Ten*; data from the entire corpus will indicate if these patterns remain prevalent throughout the entirety of the domain, or even entire selection of text.

Questions over which vocabulary lists to use in the comparison were also raised by the pilot results; while a comparison list was not run through AntConc (Anthony, 2015), analyzing the frequent terms that were present allowed for a rough comparison to the proposed vocabulary lists. The Mathematics Glossary in the Common Core is not only smaller than comparative mathematics vocabulary lists, but also spans all grade levels (*CCSS of Mathematics*, 2010). While the results of comparing the corpus of directional terms in the fourth grade textbook could be assumed to be small, this Common Core mathematical word list was kept in order to represent what classroom teachers are given as an initial resource in order to teach to the identified standards. Because of this limited comparison, the vocabulary recommended by *Math in Focus* was also included to get a better representation of which terms are typically taught in an upper elementary mathematics classroom. Finally, since the research shows the many layers of vocabulary present in the language of mathematics, including academic, content specific, and everyday, it is important to include a comparison to some of the cross-content words that are common to academic language. Zwiers (2008) created his Partial Academic Word List from Coxhead's *Academic Word List in the Academic Corpus*. Zwiers selected words that would be less obvious for instruction as they describe complex and abstract concepts across disciplines. This Partial Academic Word List was included as a final comparison (Zwiers, 2008).

Overall, the pilot raised questions on what sections of text to include, how to categorize the terms, and which word lists to use as comparison, as well as identifying the steps in using the technology to transcribe and categorize the directional sentences. By revisiting the research and identifying initial patterns present in the collected text, the questions were resolved and the results led to a more cohesive and manageable study.

Analysis of Data

The procedure was identified and tested, and the initial categorizations were set and defined before the corpus was compiled. Along with transcribing the sentences into a corpus, the identified categories were labeled in Microsoft Excel in the hopes that larger patterns would be common. A comparison was then run in AntConc (Anthony, 2015) between the textbook corpus and the three identified vocabulary lists, with the intention that an overlap of terms would be identified, or possibly a lack of, between the vocabulary lists and the corpus. Finally, the identification of frequent lexical bundles in AntConc also gave a point of analysis, particularly in comparing phrases used to terms recommended for instruction. Again, as this is a quantitative study, the patterns between and among the data sets should lead to more direction in identifying specific areas of instruction and further questions to research.

Verification of Data

While validity in transcription is almost ensured as text is copied directly from the *Math in Focus* curriculum, errors can occur on identification and labeling of linguistic features. Because of this, the data collection procedures were reviewed by a colleague experienced in linguistics. This second rater categorized sections from five chapters in the fourth grade *Math in Focus* so that one from each CCSS domain was analyzed. The results of this subset were compared to the initial study in order to set up inter-rater reliability. Any discrepancies were discussed until agreement was reached.

Conclusion

This chapter described the methods used to conduct this study on the linguistic features present in the directions of an upper elementary mathematics textbook. A quantitative analysis was used that collected the directional sentences in the fourth grade *Math in Focus* textbook into a corpus and categorized them according to linguistic features described above. The ultimate goal of this research is to identify a list of terms, lexical bundles, and linguistic features that teachers could then use to provide more explicit language instruction for ELs. Chapter Four will present the results of the analysis, which is the next step towards the identified goal.

CHAPTER FOUR: RESULTS

This study is a text analysis of directional sentences in the fourth grade *Math in Focus* textbook. Data in this study was collected to shed light on these questions: What features of the language of mathematics that researchers identify as difficult to understand are commonly used in conveying directions in an upper elementary mathematics textbook? How do the words and lexical bundles that are used for directions and the recommended vocabulary for instruction compare? After data collection and categorization, the occurrence of grammatical and syntactical features in the directional sentences was measured. The frequent words and lexical bundles were identified, while the vocabulary used for directions were also compared to the vocabulary recommended for instruction. This process indicates linguistic features that could be explicitly taught to ELs, as they are the language aspects prevalent in the curriculum.

Data Collection

Directional sentences from the fourth grade *Math in Focus* textbook were transcribed and then labeled in Excel, which allowed for multiple columns of labels and categorization. In the fourth grade *Math in Focus* textbooks, there are 1,464 directional sentences found in the various sections. This does not include sentences in the *Learn* section or sentences in word problems. The first was because, if teachers are going to provide direct instruction, they would most likely utilize the *Learn* section of text, as prescribed in the teacher's guide, therefore providing support with the language load (Fong et al., 2009). The second section was not included because previous research identifies word problems as a separate genre within the language of mathematics (Butler et al., 2004b; Dale & Cuevas, 1992; Fang & Schleppegrell, 2010; Gerofsky, 1996; Martiniello, 2008; Mestre, 2013; Moschkovich, 2012).

In transcribing the text from the remaining sections, one issue that was mediated was determining which sentences were part of a word problem, and therefore would be excluded; another issue was how to record the numbers and symbols. In the end, all sentences from the sections that were not labeled *Learn* and any sentences in the surrounding text features were recorded. As for the numbers and symbols, the text import in Excel only accepted the basic computation symbols. Since the categorization focused on language, and only needed to identify a number as one 'word' and a symbol as a separate 'word,' the symbols used did not need to be specific, for example ÷, but generally recognizable, such as /. Symbols that were a part of the number, such as indicating the name ∠ or measure ° of an angle did not have an equivalent in Excel, so the angle names and measures were entered without, such as *ABC=90*. These numerical values did not affect the analysis of linguistic patterns, as they appear as characters within a sentence.

The corpus was then run through AntConc (Anthony, 2015), which allowed for identification of the most frequent words and the most frequent lexical bundles; the most frequent words could then be compared to the everyday vocabulary in the New General Service List (Browne, 2014) to give insight into the type of vocabulary used in the corpus. AntConc also preserved a view of each word in the original position in the corpus, allowing identification of word use in context. Finally, comparing the type of vocabulary present in the corpus against the three identified vocabulary lists was an additional functional of AntConc, showing what content vocabulary in the corpus is also in the CCSS glossary (*CCSS of Mathematics,* 2010) and the fourth grade *Math in Focus* glossary (Fong et al., 2009), and what academic vocabulary is also found in Zwiers' (2008) Partial Academic Word List.

Peer Analysis

In order to ensure internal validity of the categorization of the corpus, a sample set was given to a colleague to label separately and then compare the identification. This peer reader reviewed 48 sentences, with 587 linguistic categories to mark. The two raters discussed any coding differences, using the descriptions and examples in Appendix E as the focal point, and checked that the sum of the word count in the identified verb and noun phrases equaled the total word count in the sentence. Of the data reviewed that followed the coding guidelines, the labeling was consistent at 97.7%. A particular focus on identification of relative clauses was seen as an area of need for more collaboration. Since there was a limited amount of relative clauses labeled, each selection from the corpus was discussed until 100% agreement was reached.

Linguistic Data Categorized

Once the directional sentences, location tags (such as 4.A.1.p3), section labels (such as QC for Quick Check), and CCSS identification (such as NBT for *Numbers & Operations in Base Ten*) were imported into Excel, the sentences were labeled according to the previously identified categories that researchers have found cause difficulty with comprehension:

1. Type of sentence (question, command, or statement)
2. Verb-initial sentence (yes/no)
3. Verb tense
4. Passive verb form (yes/no)
5. Number of verb phrases
6. Length of verb phrase(s)
7. Number of nominal phrases
8. Length of nominal phrase(s)
9. Length of sentence
10. Number of relative clauses

To cross-check the analysis, the length of each sentence was compared to the sum of words in the verb phrases and nominal phrases, with the goal that the numbers were equal. This equivalency further validates the categorization, since each word is either a part of a nominal phrase as the subject, object, predicate nominal, or prepositional phrase (nominal phrase), or the words are a part of a verb phrase. After labeling each sentence for the above categories, it was possible to identify the averages of each category to analyze for trends. Microsoft Excel also allowed for subcategorization within the grade level text, sorting the linguistic categories into subcategories based on their CCSS designation.

Sentence Type and Length

Part of the linguistic categorization looked at the sentence type and length. Table 1 shows one aspect of the linguistic categorization, sentence type and length. This and subsequent tables on the linguistic categories list the data in separate CCSS domains as well as the total findings in the entire corpus. All percentages were calculated using the number of directional sentences as the denominator.

Table 1

Sentence Type and Length

CCSS Domains*	Number of directional sentences	Question sentences	Command sentences	Statement sentences	Average length of sentence
OA	133	15.0%	71.4%	13.5%	7.7
NBT	293	5.8%	90.8%	3.4%	6.0
NF	331	6.0%	83.7%	10.3%	7.2
G	258	8.5%	72.9%	18.6%	8.4
MD	449	26.1%	60.8%	13.1%	9.0
Entire Corpus	1464	13.4%	75.1%	11.5%	7.8

*Operations & Algebraic Thinking (OA), Number & Operations in Base Ten (NBT), Number & Operations-Fractions (NF), Geometry (G), and Measurement & Data (MD)

Measurement & Data contains the most sentences, 449. *Operations & Algebraic Thinking* uses the least number of directional sentences at 133, while the entire corpus contains 1,464 sentences. Of the three sentence types, question, command, or statement, commands occur the most frequently, as they make up 75.1% of the total directional sentences. In the CCSS domains, *Numbers & Operations in Base Ten* contains a majority of command sentences, at 90.8%. *Measurement & Data* uses the least amount of commands, but in that domain commands still make up the majority of sentences at 60.8%. Questions appear most frequently in the *Measurement & Data* and *Operations &*

Algebraic Thinking domains, with 26.1% and 15.0% respectively. The least frequent sentence type, at 3.4%, is a statement in *Number & Operations in Base Ten*. The average length of sentences, at 7.8 words, ranged from an average of 6.0 words in *Number & Operations in Base Ten* to 9.0 words in *Measurement & Data*. Researchers identify a direct correlation between sentence length and linguistic complexity (Abedi & Lord, 2001; Bergqvist et al., 2012; Butler et al., 2004b; Martiniello, 2008).

Nominal Phrase Length and Type

The linguistic categorization of the directional sentences in the corpus identified the average number of nominal phrases, the length of each, and any relative clauses.

Table 2

Characteristics of Nominal Phrases and Relative Clauses

Nominal phrases N = 3,295			Relative clauses N = 10		
CCSS Domains*	Average number of relative clauses	Average length of nominal phrase	Percentage of relative clauses out of total nominal clauses	Number of sentences with 1 relative clause	Number of sentences with 2 relative clauses
OA	2.3	2.6	0.0%	0	0
NBT	1.6	2.5	0.0%	0	0
NF	2.1	2.6	0.1%	1	0
G	2.4	2.6	0.5%	3	0
MD	2.7	2.5	0.5%	4	1
Entire Corpus	2.3	2.6	0.3%	8	1

*Operations & Algebraic Thinking (OA), Number & Operations in Base Ten (NBT), Number & Operations-Fractions (NF), Geometry (G), and Measurement & Data (MD)

Nominal phrases are a feature of sentences that can cause confusion for language learners (Abedi & Lord, 2010; Butler et al., 2004a; Butler et al., 2004b; Haag et al., 2013; de Oliveira & Cheng, 2011; Shaftel et al., 2006; Sweeney, 2014). The number of nominal phrases in a sentence stays fairly close to the average of all sections, shown in the bottom row, which is 2.3. The exception is *Number & Operations in Base Ten*, where an average of 1.6 nominal phrases per sentence is found. As a sentence containing a subject and direct object would contain two nominal phrases per sentence, an average of 1.6 indicates that most sections range close to a sentence containing a subject and direct object. The exception is *Number & Operations in Base Ten*, which also has the highest frequency of commands, meaning that the subject is implied as the directive is to the reader. Considering this section combines the high frequency of commands and the lower average number of total nominal phrases in a sentence, it indicates that more of these sentences contain clauses that lack subjects, making them difficult to comprehend (Schleppegrell, 2007). The average length of nominal phrases also remains fairly consistent among the CCSS domains, ranging from an average of 2.5 to 2.6 words in a phrase.

Of the sentences, only 9 (0.6%) of the total 1,464 in the corpus (see Table 1) include a least one relative clause. There are 3,295 nominal phrases in the corpus, 10 (0.3%) of which are relative clauses, which can prove difficult for ELs to understand (Derewianka, 2013). *Geometry* and *Measurement & Data* contain the highest occurrences of relative clauses, 3 (0.5%) and 6 (0.5%) respectively.

Verb Phrase Length and Type

Another layer of linguistic categorization identified the length of the verb phrase and the tense and aspect of the main verb phrase. Research shows the more words involved in a verb phrase, the more complex the verb phrase becomes, as added verbs, auxiliary terms, negatives, and modals can lead to difficulty in comprehension (Abedi & Lord, 2001; Butler et al., 2004a; Fang & Schleppegrell, 2010; Irujo, 2007; Shaftel et al., 2006). Language learners are also impacted by passivation, as it can hide the doer of an action, causing confusion in agency (de Oliveira & Cheng, 2011; Shaftel et al., 2006; Spanos et al., 2013; Zwiers, 2008). These categories were analyzed, along with identification of verb-initial sentences per the high frequency of occurrence in the pilot (74%).

Table 3

Occurrence of Verb Phrase, Length, and Type

			Verb-initial sentences		Present tense sentences		Past tense sentences		Future tense sentences		Passive sentences	
CCSS Domains*	Average number of verb phrases	Average length of verb phrase	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
OA	1.2	1.5	79	59.4%	133	100.0%	0	0.0%	0	0.0%	2	1.5%
NBT	1.2	1.5	216	73.7%	289	98.6%	3	1.0%	1	0.3%	2	0.7%
NF	1.2	1.4	240	72.5%	320	96.7%	5	1.5%	5	1.5%	16	4.8%
G	1.3	1.6	181	70.2%	257	99.6%	1	0.4%	0	0.0%	4	1.6%
MD	1.3	1.6	246	54.8%	384	85.5%	63	14.0%	2	0.4%	27	6.0%
Entire Corpus	1.3	1.5	962	65.7%	1383	94.5%	72	4.9%	8	0.5%	51	3.5%

Number of verb phrases in a sentence: Min.=1 Max.=4
Verb phrase length: Min.=1 Max.=8

*Operations & Algebraic Thinking (OA), Number & Operations in Base Ten (NBT), Number & Operations-Fractions (NF), Geometry (G), and Measurement & Data (MD)

As shown in Table 3, the average number of verb phrases in a sentence in the entire corpus is 1.3, with *Measurement & Data* and *Geometry* also at 1.3. Among the CCSS domains, the range in the average number of words in a verb phrase goes from 1.4 in *Number & Operations-Fractions* to 1.6 in *Measurement & Data* and *Geometry*. The average length of the total number of verb phrases is 1.5 words; these verb phrases range in length from 1 to 8 words. The frequency of verb-initial sentences within the domains is similar to the occurrence of commands, as 79 (59.4%) and 246 (54.8%) sentences in *Operations & Algebraic Thinking* and *Measurement & Data* start with a verb. Verb initiality is still highly frequent throughout the fourth grade text, as 962 (65.7%) of the total sentences start with a verb (Appendix E).

Analyzing the various tenses of the main verb phrase, present tense is the most frequent, as 1383 (94.5%) sentences occur in present tense. Past tense is the next most common across most of the domains, ranging from 0-63 (0-14.0%) and finally future tense, ranging from 0-5 (0-1.5%) sentences. The exception is *Measurement & Data*, which has 63 (14.0%) sentences with past tense. The passive construction is also more frequent in this domain at 27 (6.0%), with *Number & Operations-Fractions* containing the next highest occurrence at 16 (4.8%). The use of passive construction, which can cause confusion for language learners by hiding the doer of the action (de Oliveira & Cheng, 2011; Shaftel et al., 2006; Spanos et al., 2013; Zwiers, 2008), is 51 (3.5%) of the directional sentences.

Identifying Commonly Used Terms

The next section of analysis comes from inputting the corpus of directional sentences into AntConc (Anthony, 2015). This program allowed for identification of the most frequent words, the most frequent lexical bundles, and a comparison between the terms in the corpus and those found in the Common Core State Standards recommended vocabulary, the vocabulary lists within the fourth grade *Math in Focus* curriculum, and Zwiers' (2008) Partial Academic Word List.

Most Frequent Words in the Corpus

The first analysis in AntConc (Anthony, 2015) looked at a list of all the words present in the corpus. AntConc identified 11,004 word tokens, or word units, and 774 word types, or different kinds of words, in the corpus. This list of words was filtered with Anthony's (2015) AntBNC Lemma List; a lemma list is a set of words that have the same meaning and are derived from a headword. *Be* is the headword for the following list: *am*, *are*, *been*, *is*, *was*, *were*. The AntBNC Lemma List filter allows for word types to be combined and analyzed under the headword, instead of derivations listed as separate words. The most 25 most frequent word types and a count of their use appear in Table 4.

Table 4

Most Frequent Word List with Lemma Form(s) from the Fourth Grade Math in Focus Textbook

Word Types N = 774			Word Tokens N = 11,004	
Rank	Word (headword)	Frequency	Lemma(s)	Grammatical Role (as used in *Math in Focus*)
1	**the***	1054		determiner
2	**of***	451		preposition
3	**a**	347		determiner
4	**to***	312		preposition
5	**be***	307	am (5), are (78), be (14), been (3), is (180), was (20), were (6)	verb
6	number	290	number (195), numbers (95)	noun, adjective
7	each	250		adjective, adverb, determiner, pronoun
8	find	212	find (209), finding (1), finds (1), found (1)	verb
9	**and***	207		conjunction
10	**you**	196	you (70), your (126)	pronoun
11	**in**	170		preposition
12	use	163	use (127), used (10), uses (2) using (24)	verb
13	line	153	line (139), lines (14)	noun
14	figure	119	figure (85), figures (34)	noun
15	**on**	118		preposition
16	answer	98	answer (69), answers (29)	noun, verb
17	draw	93	draw (70), drawing (15), drawn (4), draws (1), drew (3)	verb, adjective
18	angle	89	angle (49), angles (40)	noun
19	decimal	87	decimal (69), decimals (18)	noun
20	square	86	square (63), squares (23)	noun
21	**that**	85		determiner
22	complete	83	complete (81), completed (1), completes (1)	verb
23	show	83	show (44), showing (1), shown (13), shows (25)	verb
24	then	74		adverb
25	**these**	73		determiner, pronoun

*A word in **bold** indicates that is also appears among the top 25 most frequent words in the New General Service List (Browne, 2014), and an * means it appears among the top five.*

Of the 25 most frequent words, 11 are also in the top 25 most frequent words in the New General Service List (NGSL) (Browne, 2014), *the*, *of*, *a*, *to*, *be*, *and*, *you*, *in*, *on*, *that*, and *these*; four of the top five most frequent words from *Math in Focus* are also in the top five of the NGSL, *the*, *of*, *to*, and *be*. The most frequent words in this corpus that do not appear in the NGSL are all nouns or verbs, *number*, *find*, *use*, *line*, *figure*, *answer*, *draw*, *angle*, *decimal*, *square*, *complete*, and *show*. The exception is *each* and *then*. The grammatical role of each word was determined by viewing the word in AntConc's (Anthony, 2015) concordance, showing the term in its position in the text, and identifying its syntactical role in the sentence. For example, *drawing* is used as an adjective in the sentence: "Use a computer drawing tool to draw these figures" (Fong et al., 2009, 4B p. 136). All of the nouns and verbs in the list of 25 most frequent words are words that can appear in different grammatical roles except for *decimal*. For example, *show* can serve as a verb, *I will show you the answer,* or as a noun, *We went to the show*. While there are multiple grammatical roles for all the words except one, in this corpus only *number*, *each*, *these*, and *answer* are used in more than one grammatical role; *find*, *use*, *line*, *figure*, *angle*, *square*, *complete*, and *show* are only utilized in one grammatical role in *Math in Focus*. *Figure* is a polysemous word, but viewing all 119 occurrences in AntConc, it is only used as a noun with the meaning of *diagram* or *shape*. The other polysemous word among the 25 most frequent words in the corpus is *draw*, and, as seen in AntConc, it is used in both polysemous roles as a verb or adjective meaning *to sketch*/ *sketched* or *to pull out/ selected*.

Most Frequent Lexical Bundles in the Corpus

AntConc (Anthony, 2015) also identifies the most frequent lexical bundles, or N-Grams, with the 15 most frequent listed in Table 5 out of 24,156 types in the corpus. An N-Gram Size of 3-5 was used, as recommended by Herbel-Eisenmann et al. (2010).

Table 5

Most Frequent Lexical Bundles in the Fourth Grade Math in Focus Textbook

Rank	Frequency	Lexical Bundle (3-5 tokens)
1	35 (27)	find the missing (numbers)
2	29	the number of
3	27 (22)	(find) the area of
4	26	to check that your
5	26	to the nearest
6	21	draw a line segment
7	20	look at the
8	19	as a decimal
9	19	the number line
10	18	in simplest form
11	18	what is the
12	17	show your work
13	16	to help you
14	16	to find the
15	15	(to check that your) answers are reasonable

Several of the lexical bundles are command phrases, such as *find the missing*, *find the area of*, *draw a line segment*, *look at the*, and *show your work*. Infinitives are also common lexical bundles: *to check that your*, *to help you*, *to find the*, and *to check that your answers are reasonable*. The only question phrase, *what is the*, appears 18 times out of a total of 24,156 types of lexical bundles. The remaining lexical bundles act as nominal or adverbial phrases: *the number of*, *the area of*, *to the nearest*, *as a decimal*, *the number line*, and *in simplest form*. Ten of the fifteen most frequent lexical bundles contain a word signaling the start of a prepositional phrase (*of*, *to*, *at*, *as*, and *in*).

Terms that Occur in the Corpus and in the *Math in Focus* Recommended Vocabulary

Math in Focus provides a glossary in the back of each textbook, which combined includes 113 terms (Fong et al., 2009). This text feature for students identifies the terms recommended for explicit instruction; teachers see these same terms throughout their instructional manual, as the start of each lesson recommends explicit instruction on their use (Fong et al., 2009). When these terms first occur in the textbook, they are listed in a table at the start of the lesson, and then found in context, bolded and highlighted. Sometimes the term is also accompanied by a definition, such as "A quotient is the answer to a division problem" (Fong et al., 2009, 4A p. 101). As these terms are sometimes composed of more than one word, such as *line segment*, rather than comparing the occurrence of individual words *line* and *segment* between the corpus and the *Math in Focus* glossary, a comparison was run that preserved the 113 terms instead of comparing individual word units. This analysis also allowed for derivations, such as counting *vertical lines* when looking at the frequency of the term *vertical line*. Table 6 shows the 25 most frequent terms out of the 80 terms that appear both in the corpus and in the fourth grade *Math in Focus* Glossary.

Table 6

Most Frequent Terms that Occur in the Corpus and in the Fourth Grade Math in Focus Glossary

Rank	Frequency	Term
1	153	line(s)
2	89	angle(s)
3	88	decimal(s)
4	86	square(s)
5	72	rectangle(s)
6	61	line segment(s)
7	42	table(s)
8	41	round(s, ed, ing)
9	39	area(s)
10	37	estimate
11	37	reasonable
12	34	mixed number(s)
13	32	length
14	30	greatest
15	30	multiple(s)
16	25	data
17	25	least
18	22	improper fraction(s)
19	19	factor(s)
20	18	perimeter
21	18	right angle(s)
22	18	simplest form
23	16	order
24	15	greater than
25	15	line graph

The first four terms, *line(s)*, *angle(s)*, *decimal(s)*, and *square(s)*, also appear in the 25 most frequent word list in the corpus, as seen in Table 4 at numbers 13, 18, 19, and 20 respectively. Three of the terms are defined as verbs in the *Math in Focus* glossary, *round*, *estimate*, and *order*, and four are adjectives, *reasonable*, *greatest*, *least*, and *greater than*. The rest of the terms function as nouns in their definition. Of the 113 terms in the *Math in Focus* glossary, 80 (70.8%) appear in the directional sentences in the corpus.

Terms that Occur in the Corpus and in the CCSS Glossary

The list of recommended Common Core State Standard vocabulary is found at the end of the list of standards in the glossary as a list of 55 terms, not individual words (*CCSS of Mathematics*, 2010). Comparing the individual words with that of the corpus gives results of overlapping words such as *numbers*, *whole*, and *form*. But as the CCSS glossary gives definitions of terms, such as *expanded form*, so the comparison run through AntConc (Anthony, 2015) was re-run preserving those terms. With that filter, the concurrence between the two lists only results in four overlapping occurrences of terms out of the 55 (7.3%). These terms (and their frequency) are: *fraction(s)* (65), *whole number(s)* (10), *congruent* (5), and *expanded form* (5). *Congruent* and *expanded form* can also be found in the *Math in Focus* glossary (Fong et al., 2009).

Words that Occur in the Corpus and in Zwiers' Partial Academic Word List

The words in the corpus were then compared to the words found in Zwiers' (2008) Partial Academic Word List. This list, which is a subset of Coxhead's *Academic Word List in the Academic Corpus*, contains 227 different words; all 13 words that appear in both the corpus and in the Partial Academic Word List are in Table 7 with their frequency of use in the corpus.

Table 7

Words that Occur in the Corpus and in Zwiers' Partial Academic Word List

Rank	Frequency	Word
1	39	area
2	37	estimate
3	25	data
4	20	parallel
5	15	factors
6	7	label
7	4	method
8	2	principal
9	1	create
10	1	formula
11	1	occur
12	1	select
13	1	similar

Only 13 of the 227 words (5.7%) in Zwiers' (2008) Partial Academic Word List appear in the corpus. As Zwiers identifies abstract and complex forms of less obvious words for the purpose of instruction, his list does not include derivations such as *labeling* and *labels* when *label* is the identified word. This reduced the frequency of occurrence for some words.

Comparison between the Corpus and Three Vocabulary Lists

After comparing each of the three vocabulary lists to the corpus, the overlap of all the lists together was identified to look for a larger pattern of vocabulary use, as seen in Figure 1.

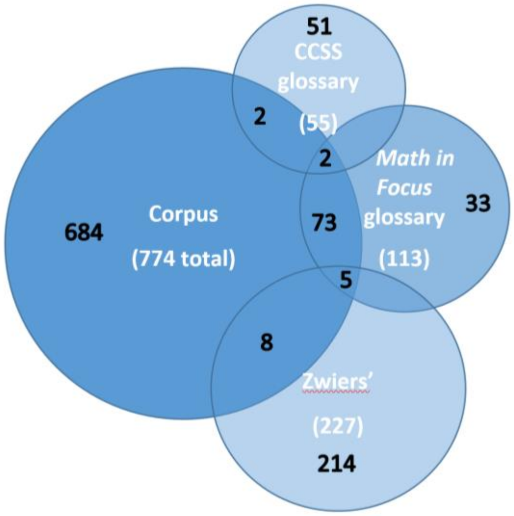

Figure 1: Terms with (total) and **count**

Five words from Zwiers' (2008) Partial Academic Word List appear in the *Math in Focus* glossary (Fong et al., 2009) and the corpus, *area*, *estimate*, *data*, *factors*, and *parallel* (in *parallel line segment(s)*), as seen in the overlapping circles in *Figure 1*. This is a larger co-occurrence than the overlap between the CCSS (*CCSS of Mathematics*, 2010) and *Math in Focus* glossary terms that also appear in the corpus, as only two, *congruent* and *expanded form*, appear in all three text selections.

Conclusion

This chapter presented the results of the data analysis of the corpus created from the fourth grade *Math in Focus* textbook. The data displays the prevalence of previously identified linguistic features as well as lists of frequent terms and lexical bundles, both in the corpus and in the overlap of previously identified vocabulary lists. The prevalent linguistic characteristics in the 1,464 directional sentences are commands at 75.1%, and average sentence length of 7.8 words, 2.3 nominal phrases per sentence with a length of

2.6 words. Only 9 (0.3%) of directional sentences contain one or more relative clause. In the directional sentences there is an average of 1.3 verb phrases with a length of 1.5 words. Verb initiality is frequent, 962 (65.7%), as is present tense, 1383 (94.5%), while the passive construction is infrequent at 51 (3.5%). The 25 most frequent words found in the corpus show 11 that also appear in the New General Service List (Browne, 2014). Almost all of the remaining frequent terms are nouns and verbs that are only utilized syntactically in one grammatical role in the directional sentences. Ten of the fifteen most frequent lexical bundles contain a word signaling a prepositional phrase. Comparing the words in the corpus to those in the three vocabulary lists, 80 (70.8%) of the 113 content terms in the *Math in Focus* glossary (Fong et al., 2009) appear in the corpus, as well as four (7.3%) of the 55 content terms in the CCSS glossary (*CCSS of Mathematics*, 2010), and 13 (5.7%) of the 227 academic words from Zwiers' (2008) Partial Academic Word List. The final chapter will discuss the major findings from these data points, their implications, and suggestions for further research.

CHAPTER FIVE: CONCLUSIONS

This linguistic analysis was conducted to answer two questions. First, what features of the language of mathematics that research has shown are difficult to comprehend are commonly used in conveying directions in an upper elementary mathematics textbook? Second, how do the words and lexical bundles that are used for directions and the recommended vocabulary for instruction compare? This final chapter will address the major findings from this study, limitations, implications for teachers of ELs, and present suggestions for further research.

Major Findings

After categorizing the data in the labeled linguistic categories and running an analysis to indicate the overlap of vocabulary, the resulting data was analyzed. The goal of analysis was to identify substantial occurrences of linguistic features and vocabulary terms in order to find a frequent language form on which ELs might benefit from explicit instruction. This language instruction could then support ELs in their comprehension and acquisition of the language of mathematics.

Possibly Difficult Linguistic Features of Directional Sentences

Sentence type and length. Within the corpus, the first area of linguistic analysis that researchers have identified as causing difficulty of comprehension is the categorization of sentence type and length. There is high frequency of commands in the corpus, 75.1% of all directional sentences; commands have a linguistic style that might be difficult for ELs to comprehend. The majority of commands are verb-initial sentences, 957 (not including the five verb-initial statement sentences), which are 94.8% of all the command sentences. This verb-initial construction can hide the point of view, therefore obstructing understanding (Schleppegrell, 2007). For example, a student might not read “Complete,” (Fong et al., 2009, 4A p. 4) as a directive to finish the problem, but instead as a label describing something as finished. If ELs are not only unfamiliar with hearing commands, but also in reading a sentence with a dropped subject, they might be unsure of where to look to find what to do. The syntactical structure of their home language might also come into play, as any direct object in a subject-implied sentence might be mistaken as the subject. For example, “Compare the numbers,” (Fong et al., 2009, 4A p. 4) might be interpreted as *the numbers compare* if a student’s first language is in any variation of the verb-subject structure; that student, or any other EL, might assume every sentence must have a subject and therefore mistakenly interpret the sentence as a statement instead of an instruction to do something.

Commands are also a linguistic style of presentation that might not be present in oral discourse in a mathematics classroom, as identified by several research studies analyzing oral discourse samples across grade levels, gender, geographic location, and socio-economic status (Butler et al., 2004b; Herbel-Eisenmann & Wagner, 2010; Herbel-Eisenmann et al., 2010). The command structure present in the *Math in Focus* textbook conveys mathematics in an authoritative format, consistent with de Oliveira and Cheng’s (2011) analysis of mathematical discourse and Butler et al.’s (2004b) identification of word problems. This can be in contrast of the typical view of mathematical texts as procedural (de Oliveira & Cheng, 2011). Herbel-Eisenmann et al. (2010) found in their analysis of oral discourse in secondary mathematics classrooms that the most common lexical bundles are stance bundles, where the teachers use phrases that conveyed personal feelings, attitudes, value judgments, or assessments, such as "I want you to," and "You don't have to" (p. 48). Stance bundles are often used by teachers to appear polite and less authoritative, but these intentions are not readily apparent to students and therefore might hide the mathematical significance of an activity. The lack of bundles conveying absolutes, unknowns, and commands in discourse in the mathematics classroom perpetuates the view of mathematics as being an abstract concept (Butler et al., 2004b; Herbel-Eisenmann & Wagner, 2010; Herbel-Eisenmann et al., 2010). The frequency of commands found in the corpus does not match the prevalence of indirect bundles as found by these studies of oral discourse, indicating a possible discrepancy between written and oral language use in the mathematics classroom.

In the written language of mathematics, sentence length affects text comprehension (Abedi & Lord, 2001; Butler et al., 2004b; Martiniello, 2008). Average length of the directional sentences in the corpus is 7.8 words. This is similar to Butler et al.'s (2004b) analysis of word problems in fifth grade mathematics textbooks at an average of 8 words in a sentence, but fewer than Butler et al.'s (2004a) finding of 10.8 words per sentence in word problems. Martiniello (2008) also analyzed word problems in a mathematics assessment and found a range of average sentence length from 8 to 16, but not a direct indication of what length of a sentence affects comprehension. Words problems are a distinct genre within the language of mathematics (Butler et al., 2004b; Dale & Cuevas, 1992; Fang & Schleppegrell, 2010; Gerofsky, 1996; Martiniello, 2008; Mestre, 2013;

Moschkovich, 2012), so identified sentence length in the research studies is not directly applicable to the context of directional sentences present in the corpus. Thus the long sentences in the corpus could indicate linguistic complexity, but this feature alone might not lead to difficulty of comprehension for ELs.

Sentence composition. The specific linguistic features identified in the sentences throughout the corpus can cause difficulty of comprehension, whether through length of verb or nominal phrases, or through specific features of the language of mathematics present therein. Martiniello's (2008) study shows multiple clauses in the sentence constructions of word problems can lead to difficulty of comprehension; the directional sentences in the fourth grade *Math in Focus* contain an average of 2.3 nominal phrases and 1.3 verb phrases. Verb phrases with three or more words are identified as complex verbs, which can include verbs, modals, and auxiliaries (Shaftel et al., 2006). The average length of verb phrases in the corpus is 1.5 words, which would indicate about half the phrases include only the head verb. The average of 2.3 nominal phrases in a directional sentence is similar to Butler et al.'s (2004b) average of 1.9 nominal phrases per sentence in word problems, but less than the 3.4 of Butler et al.'s (2004a). A sentence containing a subject and a direct object would consist of two nominal phrases, so an average of 2.3 indicates the majority of directional sentences are similar to this composition. The average length of nominal phrases in the corpus at 2.6 words is also similar to the 2.4 of Butler et al.'s (2004a) analysis of word problems.

More nominal phrases in a sentence could indicate a greater use of logical connectors and comparative structures that would join independent or dependent clauses. Logical connectors and comparative structures are another difficult feature of the language of mathematics (Butler et al., 2004b; Dale & Cuevas, 1992; Fang & Schleppegrell, 2010; Irujo, 2007). As there is not a high occurrence of nominal phrases in the corpus, it can also be assumed there is a less frequent use of logical connectors and comparative structures. Therefore, this is not a high area of linguistic complexity in the corpus.

The overall infrequency of relative clauses in the corpus composed of directional text is also in contrast to their presence within complex nominal groups found in mathematical discourse and text in word problems (Abedi & Lord, 2001; Butler et al., 2004a; de Oliveira & Cheng, 2011). Of the 3,295 nominal phrases, only 10 (0.3%) are relative clauses. Their infrequent occurrence indicates that overall this might not be a cause of linguistic complexity in the corpus.

While tense variation in word problems add to the difficulty of comprehension (Gerofsky, 1996; Irujo, 2007; de Oliveira & Cheng, 2011), the majority of directional sentences in the fourth grade *Math in Focus* are present tense, 1383 (94.5%). There is also an infrequent use of the passive construction at 51 (3.5%) sentences, similar to Butler et al.'s (2004a) finding of 4% in word problems. Combined with the high frequency of present tense, this would indicate that the directional sentences would be less linguistically complex in regards to verb tense, as passive construction and past and future tense are deemed by researchers as verb forms that cause difficulty in comprehension (Abedi & Lord, 2001; Butler et al., 2004a; Fang & Schleppegrell, 2010; Gerofsky, 1996; Irujo, 2007; de Oliveira & Cheng, 2011).

Symbols used in the language of mathematics are an ever-present contributor to difficulty in comprehension (Adams, 2003; Dale & Cuevas, 1992; Irujo, 2007; Mestre, 2013). The directional sentences transcribed in the corpus did not include the numerical problems for which the directions were given, but students are still required to interact with these mathematical symbols, whether found in equations, labels of a shape, or data in a table. The unique text organization in a mathematics textbook requires slower and more careful reading of the spatial positioning of numbers, symbols, and text; visual elements that represent mathematical content and real-word objects are other factors of linguistic complexity that are present regardless of analysis (Dale & Cuevas, 1992; Lager, 2006; Martiniello, 2008; de Oliveira & Cheng, 2011). Taken with the occurrence of nominal and verb phrases identified in other research studies (Abedi & Lord, 2010; Butler et al., 2004a; Butler et al., 2004b; Dale & Cuevas, 1992; Fang & Schleppegrell, 2010; Gerofsky, 1996; Irujo, 2007; Martiniello, 2008; de Oliveira & Cheng, 2011), these features of the language of mathematics might lead to difficulty in comprehending the fourth grade *Math in Focus* text for ELs, although this study did not analyze the impact of symbols and text features.

Vocabulary of Directional Sentences

Words and lexical bundles. The analysis of the most frequent words and lexical bundles in the corpus provides insight into the type of language students will interact with throughout the text, as well as identifying vocabulary that researchers have found problematic. Of the 25 most frequent words in *Math in Focus*, 11 are everyday vocabulary found in the top 25 most frequent words in the New General Service List (Browne, 2014) and the remaining are nouns and verbs that can occur in more than one grammatical role (except for *decimal*).

Words with multiple meanings can cause difficulty in comprehension for ELs (Adams, 2003; Butler et al., 2004b; Lager, 2006; Martiniello, 2008; Schleppegrell, 2007; Shaftel et al., 2006). *Figure* is a polysemous term that can cause confusion, particularly given its prevalence in the text at 119 occurrences. In the fourth grade *Math in Focus* it is only used as a noun meaning *diagram* or *shape*; there is no use as the verb form *figure out*, meaning *to solve* or *identify*, even though this definition might be more common across linguistic registers. Since ELs are more likely to assign the familiar use of a word (Mestre, 2013), they might view *figure* as a verb and not recognize its mathematical use as describing a shape. *Draw* (93 occurrences) is another polysemous term, and its different uses appear in *Math in Focus*, as it is used to mean *to sketch*, "Draw angles with these measures," (Fong et al., 2009, 4B p. 104) or *to pull out* or *select*, such as "Each person takes turns drawing five number cards each," (4A p. 17). Again, ELs need to be familiar with the multiple meanings of vocabulary, which could hinder application of their mathematical content knowledge (Barwell, 2005; Lager, 2006; Shaftel et al., 2006; Zwiers, 2008).

Another important pattern among the 25 most frequent words in the corpus lies in the 13 words that are not found in the New General Service List (Browne, 2014). These non-everyday words can be used in multiple grammatical roles, but only appear in one syntactical position in the fourth grade *Math in Focus*. This could cause confusion if a student is not sure which grammatical role the word has in the directional sentence. *Number*, *find*, *use*, *line*, *figure*, *angle*, *square*, and *show* can all appear as a noun or a verb, but are only utilized in one grammatical function in *Math in Focus*. *Complete* is another word that can be utilized as a verb and adjective, but only appears as a verb. Not understanding the multiple grammatical functions of words can cause confusion for ELs, and the frequent use of syntactically diverse words is an indicator of linguistic difficulty (Barwell, 2005; Dale & Cuevas, 1992; Zwiers, 2008).

The prevalence of everyday vocabulary among the 25 most frequent words in the corpus, 11 out of the 25, is also common in the mathematics register; Butler et al.'s (2004a) analysis of word problems in mathematics textbooks shows a composition of only 4.9% academic words and 6.6% content words. Research on words problems identifies that this genre can be difficult for ELs to comprehend, particularly in how word problems use everyday language in complex linguistic structures (Abedi & Lord, 2001; Adler, 1998; Bergqvist, Dyrvold, & Österholm, 2012; Haag et al., 2013; Lager, 2006; Martiniello, 2008; Shaftel, Belton-Kocher, Glassnapp, & Poggio, 2006). The combination of everyday language and academic terms in mathematics assessments is also found to be problematic for ELs to navigate (Barwell, 2005; Irujo, 2007; Martiniello, 2008; Monaghan, 1999; Rubenstein & Thompson, 2002). With the high frequency of everyday vocabulary in the corpus, this is an aspect of the language of mathematics that could cause ELs to struggle in understanding the directional sentences.

The frequent lexical bundles in the corpus are another linguistic aspect that gives insight into what language patterns students encounter when reading the *Math in Focus* textbooks. Ten of the fifteen most frequent lexical bundles contain a preposition: *of*, *to*, *at*, *as*, or *in*. They, combined with the prevalent prepositions among the top 25 most frequent words in the corpus, *of* (451 occurrences), *to* (312), *in* (170), and *on* (118), might indicate a source of linguistic difficulty. Prepositions can confuse ELs because they add another dimension to the sentence that needs to be understood (Butler et al., 2004a; Irujo, 2007; Martiniello, 2008; de Oliveira & Cheng, 2011; Shaftel et al., 2006; Spanos et al., 2013; Sweeney, 2014). Prepositional phrases were counted as nominal phrases; while the length of nominal phrases in the corpus might not indicate linguistic complexity, categorization by type of nominal phrase might provide further insight into the use of complex nominal phrases in the directional sentences.

Commonality between corpus and vocabulary lists. The corpus was compared to the content vocabulary in the Common Core Mathematics Glossary (*Common Core State Standards for Mathematics*, 2010) and the *Math in Focus* glossary (Fong et al., 2009), and to the academic vocabulary identified by Zwiers (2008) in order to determine what vocabulary is used in the directional sentences. Once identified, the specific terms and their subcategory of vocabulary, content or academic, could be useful for instructors, as they then know what to provide direction instruction on in order to support ELs in acquiring the language of mathematics. Comparing the corpus to the vocabulary lists, it is no surprise that the *Math in Focus* glossary has the most commonality at 80 terms (70.8%). The fact that all terms in the *Math in Focus* glossary are not present in the directional sentences might indicate that these terms are used by instructors to talk about content or provide instruction, but not to give direction or steps to follow. The terms in the vocabulary list might also appear in word problems, which were not included in the corpus, instead of in the directional sentences. The majority of CCSS recommended vocabulary is also not found in the directional sentences in the fourth grade *Math in Focus* textbook, as only four (7.3%) terms appear in both. This lack of use might indicate a lack of content vocabulary in the *Math in Focus* directional sentences, or it might reflect the language use of the CCSS.

When analyzing the directional language used in the *Math in Focus* corpus, there is already an indication of frequent use of everyday terms. But the use of academic terms in conveying directions is lacking; 13 (5.7%) academic words from Zwiers' (2008) Partial Academic Word List also are found in the corpus. This low use of academic vocabulary is consistent with Butler et al.'s (2004a) research of fifth grade mathematics textbooks where 4.9% of the words used in word problems are academic words that appear across content areas. Whether the selection of Zwier's (2008) Partial Academic Word List was too small (227 terms) compared to Coxhead's (2000) original *Academic Word List in the Academic Corpus* (570 word families), or *Math in Focus* does not often use academic words in directional sentences, the infrequent occurrence of academic vocabulary might indicate this is not an area of linguistic difficulty.

The more complex the grammatical and lexical features, the higher the linguistic difficulty (Haag et al., 2013). There is not a measure of how many features need to be present in a language in order for it to be labeled linguistically complex; while the language found in the directional sentences in the fourth grade *Math in Focus* textbook does not contain a high occurrence of all the linguistic features of the language of mathematics that researchers identified as difficult, there are various complex linguistic features found in the directional sentences that could indicate a difficulty of comprehension for ELs. Therefore, it is important for teachers of mathematics to also provide instruction on these linguistic features of the language of mathematics.

CCSS Domain Analysis

When the directional sentences were compiled into the corpus, they were labeled with the CCSS tags in order to analyze the linguistic variances between the domains. Previous research on an upper elementary standardized assessment indicates the mathematics content areas of data analysis, statistics, and probability contain more linguistically difficult features than other topics (Martiniello, 2008).

In the corpus, the CCSS domain of *Number & Operations in Base Ten* contains the highest frequency of commands, 90.8%, and also the lowest average number of nominal phrases, 1.6, along with the shortest sentences, 6.0. That gives this domain a higher combination of linguistic features that are indicators of linguistic difficulty that the other four domains (Abedi & Lord, 2001; Bergqvist et al., 2012; Butler et al., 2004b; Martiniello, 2008). Because *Number & Operations in Base Ten* is often considered the foundation of mathematical skills (*CCSS of Mathematics*, 2010), when instructors are focusing on these base skills they should also provide explicit instruction on linguistic features in order to support ELs understanding and acquisition of the language of mathematics.

Measurement & Data contains the longest sentences at an average of 9.0. Butler et al. (2004a) also found that *Measurement & Data* contains the longest number of words per sentence, with 11.6 in ratio word problems. This domain also has a slightly higher than average length of verb phrases, 1.6, and the most phrases on average, 1.3, the same as *Geometry*. *Measurement & Data* has 63 (14.0%) of sentences in past tense and the highest occurrence of the passive construction, 27 (6.0%), all indicators of linguistic difficulty (Abedi & Lord, 2001; Bergqvist et al., 2012; Butler et al., 2004a; Butler et al., 2004b; Fang & Schleppegrell, 2010; Irujo, 2007; Martiniello, 2008; de Oliveira & Cheng, 2011; Shaftel et al., 2006; Spanos et al., 2013; Zwiers, 2008). *Measurement & Data* has a greater occurrence of these difficult linguistic features compared to the other domains, similar to Martiniello's (2008) findings, so it is important for teachers to provide more language instruction for ELs when working with this content.

Limitations

One limitation of this study is the small amount of data analyzed. While the corpus includes 1,464 directional sentences, 11,004 word tokens, and 774 word types, this is a small section compared to other corpus studies analyzing frequent word use, such as the composition of the New General Service List which analyzed 2.5 million word tokens (Browne, 2014), the Academic Word List at 3.5 million word tokens, (Coxhead, 2000) and an analysis of lexical bundles in the discourse in mathematics classrooms with a corpus of 679,987 word tokens (Herbel-Eisenmann & Wagner, 2010; Herbel-Eisenmann et al., 2010). Another area of limitation is the categorization of the linguistic terms. Each linguistic term was explicitly described and illustrated with several examples, both in the research methodology and in Appendix E, but linguists might vary on how they would describe and identify each particular linguistic feature. This could then impact the results in the tables that show the prevalence of each linguistic feature.

Because this study looks at one textbook at one grade level, it does not take into account the range of language use among textbooks and across grade levels. The comparative vocabulary lists are also limiting, as there are 113 terms in the fourth grade *Math in Focus* glossary (Fong et al., 2009), 55 terms in the CCSS glossary (*CCSS for Mathematics*, 2010), and a subset of 227 academic words identified by Zwiers (2008) from Coxhead’s (2000) list of 570 word families. Expanding the amount of words in both the corpus and the comparative vocabulary lists would provide more insight into what mathematical language is used in directional sentences.

Another limitation of this study is determining how many features of the language of mathematics need to occur in order for a text to be labeled linguistically difficult. While all of the linguistic features identified in this study have been labeled linguistically complex by previous researchers, not one of those researchers gave a base level or numerical amount of features needed to reach linguistic complexity. Because of this, while this study can identify the amount of linguistically complex items present in a text, it does not directly indicate that ELs will have difficulty in comprehending the text. These limitations are areas that would benefit from further research.

Implications for Teachers of ELs

Benefits of Explicit Instruction

With the linguistic difficulties of the language of mathematics identified in the directional sentences in the fourth grade *Math in Focus* textbook, explicit instruction on these linguistic features may provide the support ELs need to comprehend and acquire the language of mathematics (Achugar et al., 2007; Adler, 1998). Explicit instruction is providing direct teaching of a concept or skill. The instruction focuses on the teacher as expert, providing examples and non-examples to illustrate the linguistic structures and vocabulary of the concept or skill being described. The benefit of instruction that is explicit lies in the direct focus on whatever is being taught. Students, particularly ELs, can identify what they should take away as the new learning and no longer need to filter through the many terms used in instruction (Achugar et al., 2007; Zwiers, 2008). Participating in mathematical work does not necessarily lead to the acquisition of the language of mathematics; explicit instruction can better enable students from all backgrounds to participate in mathematical discourse on an equal basis (Huang & Normandia, 2007). Lager's (2006) analysis of the difficult features of the language of mathematics identifies the benefit of instruction that focuses explicitly on the aspects of language that hinder students from understanding mathematical texts. Damhuis, Segers, and Verhoeven (2014) found that explicit instruction before and after reading allows for greater long-term vocabulary growth, while Haag et al. (2013) suggest a similar practice for the academic language features used in assessments. Instruction that focuses explicitly on lexical processing strategies, or the cognitive choices readers make when encountering an unknown word, also can support vocabulary acquisition (Fraser, 1999).

The previous examples of the linguistic features used in mathematics content mean that explicit instruction needs to occur on more than just content vocabulary. Haag et al. (2013) found that academic language features receive less frequent and systematic instruction than other content vocabulary in classrooms. A combination of explicit instruction on grammatical structures as well as classroom tasks that engage students in high-quality language use could be effective in improving language skills (Achugar et al., 2007; Haag et al., 2013; Zwiers, 2008). Research with sociolinguistic models also identifies the need to make explicit the language used to convey technical meanings of mathematics, specifically the grammatical structures, discourse patterns, and features unique to the spoken and written language of mathematics (Huang & Normandia, 2007). When teacher and student oral discourse was analyzed through several research projects, the instructors' communication was found to focus on knowledge of classification, principles, and evaluation, while students' discourse centered on description, sequence, and choice. These results identified the need for explicit instruction so that students would gain the linguistic ability demonstrated by the teachers (Huang & Normandia, 2007; de Oliveira & Cheng, 2011). All of these studies show the benefit of explicit instruction that teaches the syntax, vocabulary, and discourse styles unique to the language of mathematics. Most language resources available for use in the mathematics classroom, however, provide examples of content vocabulary, which, by themselves, are not enough for mathematical language acquisition as shown by the reviewed research (Achugar et al., 2007; Adler, 1998; Damhuis et al., 2014; Fraser, 1999; Haag et al., 2013; Huang & Normandia, 2007; de Oliveira & Cheng, 2011; Zwiers, 2008).

Language in a Mathematics Classroom

As oral discourse in mathematics classrooms might convey unknown positioning and hide the authority of mathematics (Herbel-Eisenmann & Wagner, 2010; Herbel-Eisenmann et al., 2010), it is important for instructors of ELs to assess their student-directed discourse in order to ensure the language they use reflects the language their students interact with in a mathematics textbook. The more educators are aware of the kind of authority structures they are using, the better they can reflect and adjust to structures that would lead to inclusion of students and the development of their mathematical and social agency (Herbel-Eisenmann & Wagner, 2010). Their discourse should also reflect the complexity of the language of mathematics, as students should be exposed to quality input and have the opportunity to practice mathematical discourse that focuses on meaning rather than grammatical accuracy (Adams, 2003; Barwell, 2005; Dale & Cuevas, 1992; Moschkovich, 2012).

ELs would benefit from explicit instruction in several linguistic features present in the *Math in Focus* textbook. The prevalence of commands and verb-initial sentences calls for explicit instruction around their meaning and syntax (Butler et al., 2004b; Herbel-Eisenmann & Wagner, 2010; Herbel-Eisenmann et al., 2010; de Oliveira & Cheng, 2001; Schleppegrell, 2007). Instruction around nominal phrases, particularly the meaning and use of prepositional phrases, would benefit ELs as such phrases often complicate the overall meaning of the sentence (Butler et al., 2004a; Butler et al., 2004b; Irujo, 2007; Martiniello, 2008; Shaftel et al., 2006; Spanos et al., 2013). With the frequency of everyday words, and the confusion that can result from their use with academic terms (Barwell, 2005; Butler et al., 2004a; Irujo, 2007; Monaghan, 1999; Rubenstein & Thompson, 2002), it is important for mathematics instructors to highlight for students the use of everyday words and explain the different grammatical roles or polysemous meanings these words can take in a mathematical context, in addition to explicitly teaching academic and content terms (Adams, 2003; Barwell, 2005; Dale & Cuevas, 1992; Lager, 2006; Martiniello, 2008; Moschkovich, 2012; Schleppegrell, 2007; Shaftel et al., 2006; Zwiers, 2008).

With the adjustments in instruction around language, both in being explicit and also in mimicking language patterns present in mathematical text, students would also benefit if teachers identified the different linguistic choices they make and why (Achugar et al., 2007). Teaching and utilizing metalanguage would allow users to reflect on the meaning and power of linguistic choices they and others make. Metalanguage includes terms that describe the language and literacy of instruction along with terms used to describe processes, structures or concepts of language use (Joseph, 2012). It provides a common language to discuss and analyze language, allowing users to think critically about what is communicated. For example, if a text is analyzed to identify its processes and participants, a teacher can highlight how the grammatical structure and word choice can convey the position of the key participant, such as identifying them as the agent versus the beneficiary. Metalanguage gives instructors a language resource to make explicit the meaning conveyed by the language in a certain discipline, something that would further benefit ELs in acquiring the language of mathematics (Achugar et al., 2007). Utilizing metalanguage to talk about and provide explicit instruction around the previously mentioned complex linguistic features of the language of mathematics could better support ELs in comprehending and acquiring this language.

Suggestions for Further Research

One suggestion for further research would be to repeat this study with a variety of types and grade levels of mathematics textbooks. It could be interesting to observe if the linguistic patterns identified in this study are unique to the *Math in Focus* curriculum, or even to the content covered in fourth grade. The linguistic categories could also be expanded, such as categorizing specific types of nominal phrases. Another area of research that might prove insightful would be to analyze the language used in other areas of the mathematics register: oral discourse between teachers and students, discussion across peer groups, and language present in the directional components of mathematics assessments. Staying within the *Math in Focus* curriculum (Fong et al., 2009), it could be interesting to analyze which terms from the glossary actually appear in the textbook and where they occur, both in the fourth grade text and across grade levels. A similar study to this could be conducted that includes the sentences from the fourth grade *Math in Focus* instructions and word problems in the corpus, and then analyzes the appearance and frequency of the glossary terms under the different textbook headings.

Using the results of this and subsequent studies, the next step would be to identify the impact on ELs. If they were to receive explicit instruction on the identified linguistic features of the language of mathematics, what might happen to ELs' ability to utilize the language of mathematics, as demonstrated by their oral and written performance on mathematics tasks and assessments? This would prove the real test of linguistic complexity in this linguistic analysis and provide purpose to the research, as the ultimate goal is to support ELs' successful use of the language of mathematics.

Conclusion

This linguistic analysis attempted to answer the following questions. What features of the language of mathematics identified by researchers as difficult to comprehend are commonly used in conveying directions in an upper elementary mathematics textbook? How do the words and lexical bundles that are used for directions and the recommended vocabulary for instruction compare? As shown throughout the data from the fourth grade *Math in Focus* corpus analysis, the common difficult features of the language of mathematics include a command and verb-initial sentence structure in present tense, with frequent prepositions and length of phrases and sentences similar to other academic texts. The majority of words and lexical bundles used in the directional sentences are everyday vocabulary. Comparing the corpus with content and academic word lists shows that both features are present in the directional sentences, but do not compose the majority of vocabulary used. The vocabulary used also does not reflect the entirety of vocabulary typically recommended for instruction per the *Math in Focus* glossary (Fong et al., 2009).

The purpose of these results is to assist in the identification of linguistic features that should be explicitly taught to ELs in order to help them achieve proficiency in the language of mathematics. That instruction needs to cover more than vocabulary, including the unique linguistic features of the language of mathematics and utilizing metalanguage to support students in processing the language input and output that occurs in the mathematics classroom. Explicit instruction should reflect the language used in directional sentences, and not the oral positioning that can occur that conveys mathematics as an abstract concept. While this study does not provide the final say in what to teach to ELs in order to support their acquisition of the language of mathematics, it provides indications of linguistic features that currently may not be taught and opens a new path to research on the directional sentences in the mathematics register. Overall, the language of mathematics is a unique language that necessitates explicit instruction on the linguistic features and vocabulary that make it difficult to comprehend and acquire.

REFERENCES

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

Achugar, M., Schleppegrell, M., & Oteíza, T. (2007). Engaging teachers in language analysis: A functional linguistic approach to reflective literacy. *English Teaching: Practice and Critique*. *6*(2), 8-24.

Adams, T. L. (2003). Reading mathematics: More than words can say. *The Reading Teacher*. *56*(8), 786-795.

Adler, J. (1998). A Language of teaching dilemmas: Unlocking the complex multilingual secondary mathematics classroom. *For the Learning of Mathematics. 18*(1), 24-33.

Anthony, L. (2015). AntConc (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Barwell, R. (2005). Ambiguity in the mathematics classroom. *Language and Education. 19*(2), 118-126.

Bergqvist, E., Dyrvold, A., & Österholm, M. (2012). Relating vocabulary in mathematical tasks to aspects of reading and solving. In *MADIF 8, The Eighth Mathematics Education Research Seminar* (pp. 61-70). Melbourne, Australia: SMDF.

Biber, D. & Conrad, S. (2001). Quantitative corpus-based research: Much more than bean counting. *TESOL Quarterly*. *35*(2), 331-336.

Browne, C. (2014). A new general service list: The better mousetrap we've been looking for? *Vocabulary Learning and Instruction*. *3*(1), 1-10. doi:10.7820/vli.v03.1.browne

Browne, C., Culligan, B & Phillips, J. (2013). The New General Service List. Available from http://www.newgeneralservicelist.org.

Butler, F., Bailey, A., Stevens, R. & Huang, B. (2004a). *Academic English in fifth grade math, science and social studies textbooks* (CSE Tech. Rep. No. 642). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Butler, F. A., Lord, C., Stevens, R., Borrego, M., & Bailey, A. L. (2004b). *An approach to operationalizing academic language for language test development purposes: Evidence from fifth-grade science and math* (CSE Tech. Rep. No. 626). Los Angeles CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Cocking, R. R. & Chipman, S. (1988). Conceptual issues related to mathematics achievement of language minority children. In R. R. Cocking & J. P. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 17-46). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Common Core State Standards for Mathematics. (2010). Washington, DC: National Governors Association Center for Best Practices & Council of Chief State School Officers.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Cummins, J. (1999). BICS and CALP: Clarifying the distinction. University of Toronto, Toronto, Canada. Educational Resources Information Center.

Dale, T. & Cuevas, G. (1992). Integrating mathematics and language learning. In P. Richard-Amato and M. Snow (Eds.), *The Multicultural Classroom.* (pp. 330-348). White Plains, NY: Longman Publishing.

Damhuis, C. M. P., Segers, E., & Verhoeven, L. (2014). Sustainability of breadth and depth of vocabulary after implicit versus explicit instruction in kindergarten. *International Journal of Disability, Development & Education*. *61*(3), 194-211. doi:10.1080/1034912X.2014.932562

Deen, J. & Hacquebord, H. (2002). Content-based language teaching: Language in the mathematics classroom. In M. Bax & C. J-W. Zwart (Eds.), *Reflection on language and language learning: In honour of Arthur van Essen* (pp. 221-236). Philadelphia, PA: John Benjamins Publishing Company.

Denzin, N. K. & Lincoln, Y. S. (2000). *Handbook of qualitative research*. Thousand Oaks, California: Sage Publications.

Derewianka, B. (2013). *A new grammar companion for teachers.* Marrickville, Australia: Primary English Teaching Association.

Fang, Z., & Schleppegrell, M. J. (2010). Disciplinary literacies across content areas: Supporting secondary reading through functional language analysis. *Journal of Adolescent and Adult Literacy*. *53*(7), 587-597.

Fong, H. K. (2009). The underpinning concept. *Math in Focus: Singapore Math by Marshall Cavendish.* U.S.A.: Houghton Mifflin Harcourt Publishing Company.

Fong, H. K., Ramakrishnan, C., & Gan, K. S. (2009). *Math in focus: The Singapore approach.* Student Books 4A and 4B. Singapore: Marshall Cavendish International.

Francis, D., Rivera, M., Lesaux, N., Kieffer, M. & Rivera, H (2006). *Practical guidelines for the education of English language learners: Research-based recommendations for the use of accommodations in large-scale assessments.* Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Fraser, C. A. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition. 21*(2), 225-241. doi:10.1017/S0272263199002041

Gerofsky, S. (1996). A linguistic and narrative view of word problems in mathematics education. *For the Learning of Mathematics. 16*(2), 36-45.

Haag, N., Heppt, B., Stanat, P., Kuhl, P., & Pant, H. A. (2013). Second language learners' performance in mathematics: Disentangling the effects of academic language features. *Learning & Instruction. 28*, 24-34. doi:10.1016/j.learninstruc.2013.04.001

Herbel-Eisenmann, B. & Wagner, D. (2010). Appraising lexical bundles in mathematics classroom discourse: Obligation and choice. *Educational Studies in Mathematics. 75*(1), 43-63. doi:10.1007/s10649-010-9240-y

Herbel-Eisenmann, B., Wagner, D., & Cortes, V. (2010). Lexical bundle analysis in mathematics classroom discourse: The significance of stance. *Educational Studies in Mathematics. 75*(1), 23-42. doi:10.1007/s10649-010-9253-6

Huang, A. & Normandia, B. (2007). Learning the language of mathematics: A study of student writing. *International Journal of Applied Linguistics. 17*(3), 294-318.

Irujo, S. (2007). So just what is the academic language of mathematics?. *The ELL Outlook*. *6*(3), 1-6.

Joseph, C. M. (2012). *Communication and academic vocabulary in mathematics: A content analysis of prompts eliciting written responses in two elementary mathematics textbooks.* University of South Florida, Tampa, FL. ProQuest Digital Dissertations. UMI: 3547434.

Kessler, C., Quinn, M.E., & Hayes, C. (1985). Processing mathematics in a second language: Problems for LEP children. In A. Labarca, and L. Bailey (Eds.), *Delaware Symposium on Language Studies: Vol. 7. Issues in L2: Theory as practice, practice as theory*. (pp. 151-163). Norwood, NJ: Ablex Publishing Corporation.

Lager, C. (2006). Types of mathematics-language reading interactions that unnecessarily hinder algebra learning and assessment. *Reading Psychology*. *27*(2), 165-204. doi:10.1080/02702710600642475

Mackey, A. & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum.

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

Mestre, J. P. (2013). The role of language comprehension in mathematics and problem solving. In R. R. Cocking & J. P. Mestre (Eds.), *Linguistic and cultural influences* *on learning mathematics* (pp. 201-220). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Monaghan, F. (1999). Judging a word by the company it keeps: The use of concordancing software to explore aspects of the mathematics register. *Language and Education. 13*(1), 59-70. doi:10.1080/09500789908666759

Moschkovich, J. (2012). Mathematics, the Common Core, and language: Recommendations for mathematics instruction for English learners aligned with the Common Core. *Commissioned Papers on Language and Literacy Issues in the Common Core State Standards and Next Generation Science Standards* (pp. 17-31). Palo Alto, CA: Stanford University.

de Oliveira, L. C., & Cheng, D. (2011). Language and the multisemiotic nature of mathematics. *Reading Matrix: An International Online Journal*. *11*(3), 255-268.

Rubenstein, R. N. & Thompson, D. R. (2002). Understanding and supporting children's mathematical vocabulary development. *Teaching Children Mathematics. 9*(2), 107- 112.

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

Shaftel, J., Belton-Kocher, E., Glassnapp, D. & Poggio, J. (2006). The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. *Educational Assessment*. *11*(2), 105-126. doi:10.1207/s15326977ea1102_2

Spanos, G., Rhodes, N. C., Corasaniti Dale, T., & Crandall, J. (2013). Linguistic features of mathematical problem solving: Insights and applications. In R. R. Cocking & J. P. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 22-240). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Sweeney. A. (2014). *Linguistic complexity of compare word problems in elementary mathematics.* MA Thesis. Saint Paul, MN: Hamline University.

Temple, C. & Doerr, H. (2012). Developing fluency in the mathematical register through conversation in a tenth-grade classroom. *Educational Studies in Mathematics*. *81*(3), 287-306. doi:10.1007/s10649-012-9398-6

Usiskin, Z. (1996). Mathematics as a language. In P.C. Elliott and M. J. Kenney (Eds.), *Communication in mathematics, K-12 and beyond* (pp. 231-243). Reston, VA: National Council of Teachers of Mathematics.

Which student groups are making gains?. *The Nation's Report Card*. (2013). NAEP. Retrieved from http://www.nationsreportcard.gov/reading_math_2013/#/gains-by-group

Zevenbergen, R. (2000). Cracking the code of mathematics classrooms: School success as a function of linguistic, social, and cultural background. In J. Boaler (Ed.), *Multiple perspectives on mathematic teaching and learning*. (pp. 201-224). Westport, London: Ablex Publishing.

Zwiers, J. (2008). *Building academic language.* (1st ed.). San Francisco, CA: Jossey-Bass.

APPENDICES

Appendix A – Vocabulary Difficulties and Examples

Category of Difficulty	Examples
Some words are shared by mathematics and everyday English, but they have different meanings in the two contexts.	*Right* angle versus *right* answer versus *right* hand *Reflection* as flipping over a line versus *reflection* as thinking about something *Foot* as 12 inches versus the *foot* on a leg
Some mathematical words are shared with English and have comparable meanings, but the mathematical meaning is more precise.	*Difference* as the answer to a subtraction problem versus *difference* as a general comparison *Even* as divisible by 2 versus *even* as smooth
Some mathematical terms are found only in mathematical contexts.	*Quotient, decimal, denominator, quadrilateral, isosceles*
Some words have more than one mathematical meaning.	*Round* as a circle versus to *round* a number to the tenths place *Square* as a shape versus *square* as a number times itself *Second* as a measure of time versus *second* as a location in a set of ordered items
Some words shared with other disciplines have different technical meanings in the two disciplines.	*Variable* in mathematics is a letter that represents possible numerical values, but *variable* clouds in science are a weather condition.
Some mathematical terms are homonyms with everyday English words.	*Sum* versus *some, arc* versus *ark, pi* versus *pie, graphed* versus *graft.*
Some mathematical words are related, but students may confuse their distinct meanings.	*Factor* and *multiply, hundreds* and *hundredths, numerator* and *denominator*
A single English word may translate into Spanish or another language in two different ways.	In Spanish, the table at which we eat is a *mesa,* but a mathematical table is a *tabla* (Olivares, 1996).

English spelling and usage may have irregularities.	*Four* has *u* but *forty* does not. Fraction denominators, such as *sixth, fifth, fourth*, and *third*, are like ordinal numbers, but rather than *second*, the next fraction is *half*.
Some mathematical concepts are described in more than one way.	*Skip count by threes* versus *tell the multiples of 3.* *One-quarter* versus *one-fourth*
Students may adopt an informal term as if it is a mathematical term.	*Diamond* for *rhombus, Corner* for *vertex*

(Adapted from Rubenstein & Thompson, 2002, p. 108)

Appendix B – Fourth Grade *Math in Focus* Chapters Aligned to Fourth Grade CCSS

Math in Focus Chapter Title	Aligned Common Core State Standard(s)
Book 4A	
1: Place Value of Whole Numbers	4.NBT.A.1, 4.NBT.A.2
2: Estimation and Number Theory	4.OA.B.4, 4.NBT.A.3
3: Whole Number Multiplication and Division	4.OA.A.1, 4.OA.A.2, 4.OA.A.3, 4.OA.C.5, 4.NBT.B.4, 4.NBT.B.5, 4.NBT.B.6
4: Tables and Line Graphs	4.MD.B.4
6: Fractions and Mixed Numbers	4.NF.A.1, 4.NF.A.2, 4.NF.B.3, 4.NF.B.4, 4.NF.B.5
Book 4B	
7: Decimals	4.NF.C.6, 4.NF.C.7
9: Angles	4.MD.C.5, 4.MD.C.6, 4.MD.C.7
10: Perpendicular and Parallel Line Segments	4.G.A.1
11: Squares and Rectangles	4.G.A.2
12: Area and Perimeter	4.MD.A.3
13: Symmetry	4.G.A.3

(For decoding and description of the CCSS, see *CCSS of Mathematics*, 2010).

Appendix C – Fourth Grade *Math in Focus* Text Headings and Initials

Math in Focus Text Headings	Initials
Quick Check	*QC*
Guided Practice	*GP*
Let's Practice	*LP*
Hands-On Activity	*HO*
Game	*Ga*
Let's Explore!	*LE*
Math Journal	*MJ*
Put On Your Thinking Cap!	*TC*
Chapter Review/ Test	*CR*

Appendix D – Example Section from Pilot Study

location	Heading	Text	CCSS	question (Q), command (C), or statement (S)	# of words in the sentence	sum of words in verb phrases and nominal phrases	verb initiation (yes=1, no=0)	head verb tense	Passive? (yes=1, no=0)	# of verb phrases	length of verb phrase 1	length of verb phrase 2	# of nominal phrases	length of nominal phrase 1	length of nominal phrase 2	length of nominal phrase 3	length of nominal phrase 4	length of nominal phrase 5	# of defining relative clauses	# of non-defining relative clauses
4.A.1.p3	QC	Express each number in word form.	BT	C	6	6	1	present	0	1	1		2	2	3				0	0
4.A.1.p3	QC	Express each number in standard form.	BT	C	6	6	1	present	0	1	1		2	2	3				0	0
4.A.1.p3	QC	Express each number in expanded form.	BT	C	6	6	1	present	0	1	1		2	2	3				0	0
4.A.1.p4	QC	Continue each number pattern.	BT	C	4	4	1	present	0	1	1		1	3					0	0
4.A.1.p4	QC	Count on by ones, tens, hundreds, or thousands.	BT	C	8	8	1	present	0	1	2		1	6					0	0
4.A.1.p4	QC	Complete.	BT	C	1	1	1	present	0	1	1		0						0	0
4A.1 p.4	QC	Compare the numbers.	BT	C	3	3	1	present	0	1	1		1	2					0	0
4.A.1.p4	QC	Continue or complete each number pattern.	BT	C	6	6	1	present	0	1	3		1	3					0	0
4.A.1.p4	QC	Then state the rule.	BT	C	4	4	0	present	0	1	2		1	2					0	0
4.A.1.1.p7	GP	Find the missing headings.	BT	C	4	4	1	present	0	1	1		1	3					0	0
4.A.1.1.p7	GP	Express the number in word form.	BT	C	6	6	1	present	0	1	1		2	2	3				0	0
4.A.1.1.p7	GP	Express the number in standard form.	BT	C	6	6	1	present	0	1	1		2	2	3				0	0
4.A.1.1.p8	GP	Express each number in word form.	BT	C	6	6	1	present	0	1	1		2	2	3				0	0
4.A.1.1.p8	GP	Express each number in standard form.	BT	C	6	6	1	present	0	1	1		2	2	3				0	0
4.A.1.1.p8	GP	Read the number pattern.	BT	C	4	4	1	present	0	1	1		1	3					0	0
4.A.1.1.p8	GP	Find the number that comes next.	BT	C	6	6	1	present	0	2	1	2	2	2	1				1	0
4.A.1.1.p9	LP	Look at the place-value chart.	BT	C	5	5	1	present	0	1	1		1	4					0	0

Appendix E – Definitions for Linguistic Coding of Text

Linguistic Category	Definition from Derewianka (2013) and Example from *Math in Focus* (Fong et al., 2009)
Type of sentence (question, command, or statement)	Questions are sentences that ask for information, enquire about something, or probe to learn more. They are indicated in text by the use of a question mark as ending punctuation. Commands are sentences that request something or provide instruction, advice, or suggestion. A command implies that the speaker (or writer) is asking the listener (or reader) to do something. Statements are sentences that provide information. Most written text and oral presentations are composed mainly of statements; questions and commands are used when there is effort to interact with the audience. Examples: Question: *How are they alike?* (4A p. 17). Command: *Express each number in word form* (4A p. 3). Statement: *Each color shows a place value* (4A p. 21).
Verb-initial sentence	The first word in the sentence is a verb. This form is used to give instructions or directions and the subject is implied. Example: *Find the missing headings* (4A p. 7).
Verb tense	A verb is a word that conveys doing, saying, sensing, relating, or existing. The basic tenses of past, present, and future are determined by looking at the head verb and identifying the sense of time it conveys. For the purpose of this study, aspect is not identified. Examples: Past: *Who sold the least number of tickets?* (4A p. 125). Present: *How many more peaches does he have to buy?* (4A p. 131). Future: *Your partner will check your answer* (4A p. 9).
Passive verb form	Passive verb form is when the verb shifts the emphasis to the participant by moving it to the position of the doer. Sometimes the doer is omitted, implied to be someone or something other. Example: *The first one has been done for you* (4A p. 25).

Number of verb phrases	A verb phrase is the selection of text that conveys what is happening (and where). If a sentence is compound, complex, or compound-complex, there is more than one verb phrase in the sentence. Infinitives could also be recorded as a separate verb phrase if they convey a separate action. Example: *Player 1 places the counters on the place-value chart to make a 5-digit number* (4A p. 12). *Places the counters* and *to make* are both verb phrases.
Length of verb phrase(s)	The verb phrase consists of a head verb used to convey tense, any additional terms conveying aspect, auxiliaries, modals, negatives, prepositional phrases (as the circumstance), and adverbs. Conjunctions are counted in the verb phrase if they are used to connect two or more verbs within the phrase. Each of these words is counted to determine the length of the phrase. Example: *Players may choose not to use all the counters* (4A p. 12). Phrase count: 5 (*may choose not to use*)
Number of nominal phrases	A nominal phrase is the section of text that tells who or what is participating. It can also name actions, processes, states, and notions. Example: *Express each number in word form* (4A p. 3). *Each number* and *in word form* are both nominal phrases.
Length of nominal phrase	The nominal phrase contains a head noun, or pronoun, that represents a person, place, thing, or idea. This head noun can have pre-modifiers and post-modifiers, such as articles, demonstratives, possessives, adjectives and prepositional phrases (as quantifiers, classifiers, and qualifiers). Conjunctions are counted in the nominal phrase if they are used to connect two noun phrases or clauses. Numbers are also counted as nouns. Example: *Round to the nearest 10* (4A p. 30). Phrase count: 4 (*to the nearest 10*)
Length of sentence	Each word is counted in the sentence, with contractions as one word. Example: *The table shows Ms. Frey's students' favorite colors* (4A p. 139). Sentence count: 8

Number of relative clauses	Relative clauses are a type of nominal phrase that qualifies the head noun. They can be essential to the sentence, in that they specify which thing is referred to. Example: *The player who collects the most matching cards wins!* (4B p. 46). *Who collects the most matching cards* is a relative clause because it specifies which player wins.