4-2011

Investigating the World of Mathematics to Uncover How Language Proficiency Influences English Language Learners Performance on High Stakes Tests (Thesis)

Deborah Elizabeth Schuitema
Grand Valley State University

Investigating the World of Mathematics to Uncover How
Language Proficiency Influences English Language
Learners Performance on High Stakes Tests

by
Deborah Elizabeth Schuitema

April 2011

Master's Thesis
Submitted to the
Graduate Faculty of the College of Education
At Grand Valley State University
In partial fulfillment of the
Degree of Master of Education

Grand Valley State University

Acknowledgments

I would like to thank Dr. Nagnon Diarrassouba for his encouragement, advice, and constructive feedback. I benefited greatly from his help in writing this master's thesis. My thanks also go to Dr. Sango Otieno, Jake Boehmer, and Rachelle Timmer for their guidance and support with my statistical analyses. Last, but not least, I would like to thank my husband and the rest of my family for their encouragement, patience, help and support.

Deborah Elizabeth Schuitema

Abstract

English language learners (ELLs) must take the mathematics portion of state standardized tests regardless of their time spent in U.S. schools. This practice follows the misconception that mathematics is a 'universal language' and less language dependant, however, a significant performance gap between non-ELLs and ELLs on high stakes mathematics tests persists and must be addressed. In order to investigate the impact of language proficiency on high stakes test performance a cross-sectional study was conducted. The study included item performance data, by group, for 24,693 seventh and eighth grade students who took the 2007 and/or the 2008 mathematics Michigan Education Assessment Program (MEAP) assessment, and a questionnaire completed by 16 seventh and eighth grade participants for triangulation. The item performance data set was analyzed using a logistic regression model to determine the interaction effects between ELLs and non-ELLs based on item type, item language, and item strand. Cross tabulation, content, descriptive, and frequency analyses were conducted on the questionnaire responses. Findings from the logistic regression analyses show that the ratio in the odds of passing an item for ELLs and non-ELLs is affected by both whether that item was a computation or word problem, and also if a non-linguistic feature was present with $p < .0001$. The difference in passing rate for non-ELLs and ELLs was not affected by the GLCE Strand. The majority of the 19 words identified as confusing or unfamiliar on the questionnaires were context-specific or technical mathematics language features, only one of which was circled by an ELL. Results from this study have important implications for classroom instruction, test design and score interpretation.

Table of Contents

List of Tables and Figures

Chapter One: Introduction

Problem Statement

English language learners (ELLs) face the challenge of not only learning a new language, but also mastering the content for their grade level dictated by state standards and benchmarks in the same, or shorter, time frame as native speakers of English (NSEs). The pressure to acquire proficiency quickly has increased with the No Child Left Behind Act (NCLB) of 2001, requiring all students to be at or above proficiency on state standardized tests by 2014. Schools that fail to make adequate yearly progress (AYP) toward that goal face the possibility of closure, reduction in staff, or state takeover. Newly arrived ELLs do not have to take the English language arts portion of the state test until they have been in U.S. schools for one year, but they must still take the math test (Michigan Department of Education [MDE], 2010; Wright & Li, 2008) even though research shows that it takes an average of 5 to 7 years to acquire academic language or CALP (Cummins, 1981, 2000). The score from the math test will not count for the school's AYP, but this practice follows the commonly held misconception that since math is a "universal language" of numbers it will not pose great difficulty to an ELL.

Research has shown that mathematics has its own complex register including academic vocabulary, symbols and numbers (Brown, 2005; Lager, 2006; Wright & Li, 2008). Many of the items on high stakes math tests include word problems, some of which are written at a reading level above that of the students the test is designed to assess (Carter & Dean, 2006). Interpreting the scores of ELLs on math achievement tests becomes problematic because the tests are assessing both language proficiency and math skills, and the score cannot differentiate between the two (Martiniello, 2008). The test loses construct validity for this subgroup since it is assessing something that it is not intending to measure.

Importance of the Problem and Rationale for the Study

High stakes testing for adequate yearly progress has drawn great attention to ELLs and other subgroups. One seemingly positive result of including ELLs in high stakes testing is that standards have been raised for these groups and schools are being held accountable for meeting their needs (Crawford, 2004; Haycock, 2006; Kieffer, Lesaux, Rivera & Francis, 2009; Menken, 2008), however, since the consequences of falling short of AYP are so great, the washback from the test is inadvertently causing harm to the education of ELLs (Wright & Li, 2008).

Schools with high populations of ELLs will continue to suffer under this legislation because the subgroup of ELLs, by definition, will be low performing. As soon as an ELL is declared proficient in English, they are no longer labeled an ELL and their achievement scores will be combined with those of native speakers of English. Newcomers that are not proficient in English will take the place of those students that have just exited out of the subgroup, meaning that all students in the subgroup will be held accountable for achievement on standardized tests written and normed for native English speakers that have been educated in U.S. schools (Menken, 2008; Crawford, 2004). Schools are then labeled “failing” without taking into consideration language barriers for ELLs (Crawford, 2009).

Particular attention needs to be paid to the language used in high stakes mathematics tests. If the aspects of language that pose problems to ELLs are neglected, then misconceptions about mathematics and ELLs' achievements are likely to continue. It is because careful analysis has not been provided to the issue of language in mathematics that myths continue about its universality. Unless more research is conducted on the impact of language on achievement in high stakes tests and the results are used to change current procedures, it will be impossible to get an accurate picture of what an ELL knows and can do in the mathematics content area.

Background of Study

Studies have shown that ELLs continue to be among the lowest-scoring groups on standardized tests and the gap between ELLs and NSEs persists (Abedi, 2002, 2004; Martiniello, 2008; Ockey, 2007; Willner, Rivera & Acosta, 2009). Research has taken into account many different factors that contribute to this continued discrepancy including unequal access to curriculum, teacher training, language complexity on test items and accommodations.

Until the passing of the NCLB act in the last decade, ELLs were not included in standardized high stakes tests until they had achieved proficiency in the English language (Abedi, Hofstetter, & Lord, 2004; Echevarria, Powers & Short, 2006). Schools traditionally excluded ELLs from standardized testing because educators were concerned that the assessments would not provide accurate results due to language barriers and the fact that ELLs were not always given access to the same curriculum as native speakers of English (Kieffer et al., 2009; Rivera & Collum, 2004). In order to address the latter issue, standards based reform began around 1989 to promote high expectations and provide a framework for curriculum and instruction for all students.

States developed standards for the content areas so that all students would be given equal opportunity for high achievement. Following suit, the Teachers of English to Speakers of Other Languages (1997) developed a set of standards for ELLs that focused on the academic, social, and cultural language features of English. Although these standards were available, the content area standards did not make accommodations for ELLs and the majority of content area teachers have not received training in meeting the needs of ELLs in their classroom (Echevarria, Powers & Short, 2006).

According to Echevarria et al. (2006), there is a mismatch between ELLs needs and teacher preparation. Many ELLs are placed in mainstream content area classes regardless of their level of English proficiency. The issue then is that NCLB mandates that teachers must be highly qualified in their content area, but does not make similar demands for teachers of ELLs to have adequate knowledge or training in teaching methods and strategies for working with ELLs. Coates (as cited in Beal, Adams & Cohen, 2010) found that only about 15% of high school math teachers have received training in instructing ELLs.

The combination of the lack of teacher preparedness for instructing ELLs and the commonly held belief that mathematics is a "universal" language increases the language obstacles ELLs must overcome in order to achieve in mathematics classes. Strict attention must be paid to the fact that mathematics is just as language based as any other content area and that the mathematics register includes types of language to convey mathematical concepts, mathematic-specific meanings for words that are polysemic (i.e. table, one, face), and context specific vocabulary (Barbu & Beal, 2010; Brown, Cady & Taylor, 2009; Lager, 2006)

It is the linguistic complexity of word problems on high stakes mathematics achievement tests that most of the research has focused on. The majority of studies have found that linguistic complexity does affect the difficulty of math word problems for ELLs with item length having the strongest correlation (Abedi & Lord, 2001; Martiniello, 2008; Wright & Li, 2008). In an attempt to make the playing field more level for ELLs on high stakes tests with such complex language the NCLB modified legislation:

> Under Title I of the ESEA, States must include LEP students in their assessments of academic achievement in reading/language arts and mathematics, and must provide LEP students with appropriate accommodations including, to the extent practicable, assessments in the language and form most likely to yield accurate data on what LEP students know and can do in the academic content areas until they have achieved English language proficiency (U.S. Department of Education, Office of Elementary and Secondary Education, 2007, p. 3)

It is uncertain whether or not these accommodations are effective and yield valid results. A recent meta-analysis on research pertaining to the effectiveness and validity of accommodations for ELLs showed that the accommodations were not particularly effective in reducing the impact of language proficiency on high stakes test items specifically in math and science (Kieffer et al., 2009).

Since ELLs are no longer afforded time to achieve proficiency in English before joining the mainstream content area classrooms and are being held accountable on high stakes tests, considerable research and reform is still needed in regard to the impact of language proficiency on achievement on high stakes math achievement tests. The stakes are too high for students and schools for this problem to be ignored any longer.

Statement of Purpose

This study will investigate the extent to which language proficiency influences achievement on mathematics high stakes tests. The study will compare scores between English language learners (ELLs) and non-ELLs (native English speakers and former ELLs) on word problems and computation problems from standardized math assessments, and will collect data regarding confusing or unfamiliar language features in high stakes mathematics assessment items using a questionnaire.

Results from the study will inform and guide educators, administrators, legislators, test designers and parents in the assessment design, classroom instruction, and decision making that affect ELLs. As debate around NCLB continues, more information pointing to the harm high stakes testing inflicts on schools will encourage change.

Research Questions

The central research question is formulated as follows: To what extent does language proficiency influence achievement on high stakes mathematics tests? Since this question is so broad and difficult to measure, it is broken down into the following sub-questions:

1. Do English language learners perform differently on computation questions as opposed to word problems on high stakes mathematics tests? How does their achievement compare to non-English language learners on these tests?

2. Is the difficulty of an item, due to language complexity, alleviated with the inclusion of graphical, pictorial, or schematic representations?

3. Are there specific strands, as defined by the Michigan Curriculum Framework (Data & Probability, Algebra, Number & Operations, Geometry, Measurement), that English language learners perform differently on in comparison to non-English language learners on high stakes test items?

4. What language issues do students identify as confusing or unfamiliar in high stakes math test items?

Hypotheses

1. Language proficiency does influence achievement on high stakes math tests.
2. ELLs will show higher achievement on math questions that do not include many language items than on word problems.
3. Language impact on item performance will decrease with the inclusion of graphical, pictorial, or schematic representations
4. ELLs will have lower achievement on test items from the Data and Probability Strand.

Design, Data Collection and Analysis

This study utilizes a cross-sectional research design with quantitative research methods in order to better understand the influence of language proficiency on high stakes mathematics test performance. The study includes a statistical analysis of item performance by group from four different standardized tests using logistic regression models. Additionally, a questionnaire will be given to participants. Cross tabulation, frequency and content analyses will be conducted on the responses in order to provide triangulation and yield a better understanding of the issues in the study.

All of the test item performance data will be from 20 public school districts that are serviced within an Intermediate School District (ISD) in West Michigan. The questionnaires are to be given at a school located within the aforementioned school district. The questionnaire is created solely for this research study and will be piloted with a similar population sample. Before the questionnaire can be administered, permission must be granted by the IRB at Grand Valley State University and parents must sign an IRB consent form to allow their child to participate.

Definition of Terms

Accommodation – As defined by Willner, Rivera & Acosta (2009) involves changes to testing procedures, testing materials, or the testing situation in order to allow students meaningful participation in an assessment.

BICS – Basic Interpersonal Communication Skills. BICS was an acronym coined by Jim Cummins (1981) to describe the language skills needed in social situations.

CALP – Cognitive Academic Language Proficiency. CALP was an acronym coined by Jim Cummins (1981) to describe listening, reading, speaking and writing about subject area content material.

Computation Problem or Item – For the purpose of this study, a computation problem is considered a mathematical question or statement presented in words and graphics which can be solved without translating the words into mathematical or algebraic symbols. (i.e. *Solve for x, Evaluate, Determine the product of* 4×3). This definition is expanded to include definition and identification problems (i.e. *Which of the following is a square? What is the coordinate of D)* which are neither word problems nor computation problems in order to distinguish a dichotomous variable: word problem and non-word problems.

Differential Item Functioning (DIF) – As described by Young (2009) ,DIF refers to a situation where the probability of answering an item correctly differs for two groups, who have been matched on ability.

High stakes tests – As defined by Menken (2008), as tests that have important consequences for individuals and institutions.

Washback – As defined by Wall and Alderson (1993), washback is the influence of testing on teaching and learning. In other words, a high stakes test will alter the instruction a student receives in the classroom the majority of which will focus on whatever the test is assessing, therefore the test determines the classroom instruction.

Word Problem or Item – For the purpose of this study a word problem is considered a mathematical question presented in words and graphics which must be translated into mathematical and/or algebraic symbols in order to correctly determine the solution to the question (i.e. *A classroom of boys and girls contains 30 students. If the number of boys is 4 more than twice the number of girls, how many of each gender are in the class?).*

Delimitations of the Study

This thesis focuses on the relationship between language proficiency and performance on high stakes mathematics tests. It will compare the performance of middle school ELLs and Non-ELLs on computation and word problems. The study will further investigate the issue using questionnaire responses for triangulation.

The test item performance data will come from the scores of 7^{th} and 8^{th} grade students who took the mathematics portion of the Michigan Education Assessment Program (MEAP) tests in 2007 and 2008. All of the students attended schools contained in the same county Intermediate School District (ISD) in West Michigan. Although the sample from the data set will not be randomized, it includes the vast majority of the 7th and 8th graders in the population. The amount and scope of the data allow for the results to be generalizable to that specific county and any others with similar demographics in the state of Michigan.

Although ELLs differ in their level of English proficiency within-group analyses will not be conducted because of limitations in the data. The study will only focus on comparison between ELLs and non-ELLs regardless of native language, gender, or socio-economic status because this information was also unavailable in the large data set.

Furthermore, this thesis will only address mathematics assessment items from the MEAP which will limit the generalizability to counties in the state of Michigan. The MEAP assessment is designed for students in Michigan and is based on state-specific curriculum. A study of other state assessments is beyond the reach of this study due to time, resources and money.

Limitations of the Study

Much of the research to date on ELLs and math achievement on high stakes tests has focused on specific linguistic features that pose difficulty, differential item functioning, effectiveness of accommodations, and instructional strategies. The majority of this research has taken place in regions of the country that were in debate over bilingual education or that have taken specific measures to gain an accurate picture of ELLs proficiency on standardized math tests, such as creating alternate testing forms with simplified language features.

Minimal, if any, research has been done in the state of Michigan and more specifically on the mathematics assessment portion of the Michigan Educational Assessment Program (MEAP) in regard to language complexity. It would be ideal to study a wide range of students from the entire state over a long period of time, however, the access to information, timeline, and budget of the study is very limited which are also the reasons for choosing a cross-sectional design. This design is limited in that it only gives a snapshot of a point in time. Data collected at a different time period may provide dissimilar results. Moreover, causal inferences are difficult to make based on data gathered at one time.

The test item performance data analysis could be affected by several different factors. It does not account for the level of English proficiency other than students being either proficient or not proficient in English. Students that were formerly not proficient in English, but still speak English as another language, may still not have acquired all of the academic language necessary for proficiency on standardized tests. Other factors may also influence the scores such as learning disabilities, emotional or behavioral impairments, changing of standard or question type from year to year, and even test administration.

A limitation of the questionnaire is that it includes publicly released assessment items (MDE, 2007,2008) that teachers routinely use for test preparation. Students may have already experienced these specific items. Furthermore, a student must get parental consent and bring the signed form back in order to participate in the questionnaire. The response data will be limited only to those students that return the consent forms and may not be a representative sample of the population.

Lastly, my own personal beliefs and values drove the selection of the research topic, design, and method. I am also an employee at the questionnaire research site located in the aforementioned county ISD.

Organization of Thesis

The rest of this thesis is organized as follows: Chapter 2 provides a comprehensive review of the important literature related to the language complexity high stakes math tests. The research design including descriptions of the research site, the subjects involved, and the instruments and procedures used for gathering and analyzing the data are all discussed in Chapter 3. The findings from the analyses are provided in Chapter 4. Lastly, Chapter 5 presents the conclusions that can be drawn and the implications of these findings on policy and practice. Recommendations for further research are also included.

Chapter Two: Literature Review

Introduction

The inclusion of ELLs in high stakes standardized testing becomes problematic when one considers the fundamental function that language proficiency performs in the acquisition and assessment of content knowledge. Tests in content areas such as mathematics or science to some extent are also tests of language proficiency (Abedi, Courtney, Leon, Kao, & Azzam, 2006; Kieffer et al., 2009). This issue makes it difficult to make valid inferences about the achievement of ELLs. The fact that ELLs must take the state standardized mathematics test but may opt out of the English Language Arts test in the first year of attendance in U.S. schools is a reflection of the accepted notion that mathematics is less language dependent than other core areas. One tool for states to address this quandary is by providing accommodations to lessen the complexity of construct-irrelevant language, however, many of these accommodations tend to be ineffective or have questionable validity (Kieffer et al., 2009).

In order to better understand these issues, a review of the important literature relating to this topic will ensue following a discussion of the theoretical framework this research is based on. The literature falls into three general categories and will be examined in this order: 1) the language complexity of the math register, 2) the language impact on high stakes test performance of ELLs, and 3) accommodations used for ELLs to alleviate the language impact. Finally, a summary of the reviewed literature will be provided followed by a conclusion that identifies the gaps in the research that this study will address.

Theoretical Framework

The theoretical framework for this study is based on the socio-cultural theory of Vygotsky, more specifically his concept of the Zone of Proximal Development (ZPD). Vygotsky (1978), describes the ZPD as “the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance, or in collaboration with more capable peers" (pp. 86). In other words, learning occurs as a result of social interaction. The more ‘capable peer’ or adult serves as a bridge (or scaffold) between a learner’s current developmental level and their potential development. In this framework, language is not an isolated event that can be understood or acquired out of its social context.

Cummins (1981) distinction between basic interpersonal communicative skill (BICS) and cognitive academic language proficiency (CALP) relates to Vygotsky’s theory in that for the acquisition of CALP to occur it must be accessed within the learner’s zone of proximal development and scaffolded through social interaction. Cummins (2000) further elaborates on this concept:

> If students have not developed sufficient access to academic registers in either of their two languages, and if instruction does not provide the support that students need to develop this access, then their academic, linguistic, and cognitive development will not be stimulated through their classroom interactions (p. 106).

Through this quote Cummins clearly suggests that students require comprehensible input in order to further develop their learning. This concept of comprehensible input has given rise to a number of teaching and learning strategies for content area classrooms.

Despite the fact that this well founded theory is supported by empirical evidence, as shown below in the review of the literature, a number of researchers have found the theory to be controversial and have criticized Cummins' conversational and academic distinction. One of those critics, MacSwan (2000) argues that the notion of CALP represents a 'deficit theory.' This critique is founded in Chomsky's theoretical perspective that language development is complete before the age of five, thereby minimizing social and environmental influence and specifically the role that schools play in language development. MacSwan completely ignores the fact that our lexical knowledge continues to expand and that schools play a role in that expansion (Cummins, 2000).

Synthesis of Research Literature

The review of the important literature in this section will begin with a description of the language complexity of the mathematics register and then followed by an in depth look at studies examining the impact of language proficiency on high stakes test performance and the accommodations used to alleviate the language impact.

Language complexity of the mathematics register. Until recently, it was thought that mathematics learning is less language dependent than other core areas (Hansen-Thomas, 2009; Janzen, 2008; Lager, 2006) because it is more about numbers and symbols. Due to the state accountability mandates required for the sub-group of ELLs in the NCLB (2001) act, the attention of mathematics educators has started to shift toward the language needs of students in the math classroom (Lager, 2006). As a result, an increasing number of educators have focused on the content-specific vocabulary of mathematics; but since mathematics is a functional meaning making process (Lemke, 2003), more attention should be paid to engagement of students in the mathematics register through reading, writing and speaking rather than just rote memorization. This section will discuss the mathematics register, and some of the related language representations and features that have potential to confuse ELLs

The mathematics register. Cuevas (1984) defines the math register as "the meanings belonging to the natural language used in mathematics" (p.136). However, the language of mathematics contains more than just natural language (Martiniello, 2009), and this definition does not take into account that learning takes place socially and culturally (Moschkovich, 2007). O'Halloran (2000) expands this definition from the systemic linguistic functional perspective that "mathematical discourse is multisemiotic because it involves the use of the semiotic resources of mathematical symbolism, visual display and language" (p. 359). Lemke (2003) further described this perspective:

> Mathematics cannot be identified by the use of specialized mathematical symbolisms or any unique type of signs. Mathematics can be identified by the kinds of meanings it makes: meanings about addition, subtraction, multiplication, and division; about numerical difference and equality; about geometrical relationships of parallelism, orthogonality, similarity, congruence, tangency, etc., and many more in mathematical history. It is distinguished by these kinds of meanings, whether they are made by writing natural language, by drawing diagrams, or by formulating symbolic expressions (p. 1).

Learners actively negotiate these multiple meanings through a social and cultural lens as part of what this study considers as the mathematics register. Only a fraction of the mathematics register is included in the math items on standardized assessments, which currently do not take into account a socio-cultural perspective of learning (Solano-Flores & Trumbull, 2003). The learner must negotiate meaning among the different linguistic and non-linguistic representations in the context-reduced text without any form of feedback.

This process is further described by Martiniello (2009). First one must decode the problem's natural language which she refers to as "both the nonacademic or everyday language learned at home and other informal settings, and also the general cross-disciplinary academic language learned at school" (p. 162). Next, one must understand the content-specific terminology which includes vocabulary (i.e. *variable, sum, denominator)* and syntax (6 *divided by* 3, 6 *decreased by* x *is equal to 12*). By the same token, one must decode the nonlinguistic representations which include symbols (at times having their own syntax structures, i.e. $x > y - 8$), graphs, diagrams, tables and other visual representations.

Language features of mathematics that have the potential to confuse ELLs.

English language learners face the challenge of switching and translating between the mathematics register, the general English language register, and that of their native language to successfully solve mathematics problems (Lager, 2006). This challenge is further exacerbated by the fact that the algorithms and mathematical symbols used in their home country may differ from the ones used in U.S. schools (Wright & Li, 2008).

Brown, Cady, and Taylor (2009) cited many instances in mathematical language which may initially cause confusion for ELLs. A date in the U.S. is written as month/day/year (2/28/2011), but in Mexico it is written day/month/year

(28/2/2011). The majority of Europe and South America use a decimal comma to denote place value (13,8 would be read as thirteen and eight tenths) whereas North and Central America, Australia, and parts of Asia use a decimal to denote place value (13.8). Additionally, the U.S. is the only industrialized nation to use customary units of measurement (feet, yards, miles) as opposed to the metric system.

Researchers identified several difficulties ELLs may experience with mathematics textbooks and course materials. Mathematics textbooks commonly require both left-to-right and up-and-down eye movement, interpretation of non-linguistic representations, and must be read more slowly, perhaps multiple times, to develop comprehension (Wright & Li, 2008). Fillmore & Valdez (as cited in Lager, 2006) explain that ELLs may have trouble with written mathematics because meaning must be made from the language expressions, the order that they appear in and how they interconnect to be coherent.

Wright and Li (2008) conducted a linguistic analysis on released tests items from the 5th grade mathematics Texas Assessment of Knowledge and Skills (TAKS) and student worksheets for two ELLs in a 5th grade math class in Texas. The words from the assessment and the worksheets were compiled and compared using software and further analyses were conducted on both texts at the sentence level for syntactical complexity. As a result, Wright and Li (2008) found that the language complexity of the items on the 5th grade Math Texas Assessment of Knowledge and Skills (TAKS) far exceeded that of the math worksheets provided to two Cambodian students.

Lager (2006), investigated the mathematics-language reading interactions that influence learning in algebra. The sample consisted of 221 sixth and eighth grade students, 133 of which were ELLs, who participated in the study by responding to nine items specifically regarding linear patterns. All of the items included text and visual patterns. Participants were asked to highlight any unknown or confusing words phrases. Cross tabulation of responses, content analyses of the work shown, and follow-up interviews with participants were also conducted to gain a better understanding of the issue. Results showed that there are language challenges in mathematics that can hinder both ELLs and non-ELLs.

Overall, non-ELLs performed better than ELLs with a mean difference of 1.3 or 0.6 standard deviations of difference in scores, yielding a medium effect size. Responses from over half of the participants that answered the first question incorrectly indicated that the most perplexing words were *previous, extension,* and *pattern.* All three words were considered troublesome for ELLs, while *extension* was the word of main concern for non-ELLs. Misunderstandings surrounding the polysemous phrase “Figure number (N)” caused incorrect responses from 25% of the participants regardless of language proficiency or grade level. The use of variables and parentheses were also cause for confusion. In addition to language challenges explicitly identified by students, it is equally important to examine the words that students chose not to highlight. After an analysis of their written responses and interview data there was evidence that students did not correctly comprehend some of the words and phrases that were not highlighted. Lager (2006) refers to this as “false knowing” and indicates concern that students are trying to develop a complex math register on a flawed foundation.

Research has also shown that a student's level of reading proficiency can be a strong indicator of mathematical success. Beal, Adams, and Cohen (2010) investigated the correlation between reading proficiency and mathematics achievement of 442 ninth grade Algebra 1 students in four Los Angeles, California high schools. Data sources included scores from state math tests and pre- and posttests developed by the researchers, progress reports from an online math tutorial program, self-report assessments on mathematics self-concept, and English conversational and reading proficiency levels for the 209 ELLs included in the study. The researchers found that math performance increased with English reading proficiency in a non-linear manner and that there may be a minimal reading level at which math performance will improve. Furthermore, they observed that reading proficiency was significantly related to math performance whereas speaking or listening proficiency was not.

Similarly, Lamb (2010) found that elementary and middle school students performed significantly worse on mathematics assessment items from the Texas Assessment of Knowledge and Skills (TAKS) that had a readability level that was above the student grade level.

As shown above, the complexity of the mathematics register can be problematic not only for ELLs but also non-ELLs. Yet, the situation is intensified for ELLs, as evidenced by the achievement gap in high stakes test scores for they must learn to translate between their native language, the general English register and the mathematics register. The research points to a need for more studies to focus on the facilitation of the mathematics register in the classroom, preparation for mathematics teachers regarding language, and the influence of previous schooling on the development of an ELL's mathematics register. The majority of these studies did not take into account the amount of schooling an ELL had in the U.S. The study by Wright and Li (2008) did address this issue but their research only included two ELLs from Cambodia which makes it difficult to generalize findings to a larger population. Furthermore, none of the studies take place in the Midwest which has seen recent surges in ELL populations. It is also imperative to look more specifically at how the linguistic complexity of items on high stakes mathematics tests relate to student performance.

Linguistic complexity and performance on math test items. In the past decade a number of researchers started to focus on the linguistic complexities of mathematics test items that could influence performance for ELLs (Abedi, Bailey, Butler, Castellon-Wellington, Leon, & Mirocha, 2005; Abedi & Lord, 2001; Lager, 2006; Martiniello, 2008, 2009; Ockey, 2007; Shaftel, Belton-Kocher, Glasnapp & Poggio, 2006; Wolf & Leon, 2009). They wanted to know if complexity or simplicity of language correlates with students' achievement on tests. Although there were differences among researchers, most of them found that language complexity strongly correlated with students' performance. However, the specific language features that influenced performance varied by test and grade level.

Ambiguous wording, item length, difficult vocabulary, syntactic complexity with longer sentences, and comparison problems are among some of the language features that affected performance (Shaftel et al., 2006). Most experts found that item length was the only language feature to have a consistent negative effect on item performance by ELLs (Martiniello, 2009). A handful of these studies and their results are described below.

Shaftel et al. (2006) analyzed 594 test items from the Kansas general mathematics assessment given at grades 4, 7, and 10 to examine the relationship between linguistic complexity and student performance. Through multiple regression analyses the researchers found that no pattern of item linguistic characteristics impacted item difficulties across groups. Difficult mathematics vocabulary had a consistent effect for all groups at every grade level; though, the researchers noted that this is a construct-relevant feature and should be expected to relate to item difficulty.

Likewise, two other studies conducted by Martiniello (2008, 2009) investigated the linguistic complexity of items that showed increased difficulty for ELLs as compared to Non-ELLs with equivalent math proficiency. Both studies utilized results from differential item functioning (DIF) procedures conducted on the 2003 Massachusetts Comprehensive Assessment System 4th grade mathematics test. Ten out of the 39 publicly released items from the MCAS showed DIF disfavoring ELLs (Martiniello, 2008, 2009). Think-aloud protocols administered to 24 ELL fourth graders in the 2008 study confirmed that the linguistic complexity of those 10 items was most likely the cause of DIF. Some of the shared linguistic features found in those 10 items include: multiple clauses, long noun phrases, unfamiliar vocabulary, polysemous words, and words or expressions referencing mainstream American culture (i.e. *coupon, spelling bee championship)* (pp. 358).

In addition to linguistic complexity, Martiniello (2008) found that the curriculum learning strand of Data Analysis, Statistics, and Probability was also a likely source of DIF disfavoring ELLs. Five out of the seven items assessing this strand were flagged as items that potentially disfavor ELLs. The researcher gives two possible explanations for this finding: either ELLs had less exposure to the curriculum content of this learning strand, or those items involved greater semantic and linguistic complexity. Nonetheless, she suggests that more research be conducted on the relationship of this strand and performance by ELLs.

Additionally, Martiniello (2009) explored the relationship between non-linguistic forms of representation (either schematic or pictorial) and DIF measures. Results showed that the impact of linguistic complexity was weakened when an item included a nonlinguistic schematic representation. Schematic representations were described as those representing spatial or mathematical relationships such as equations, diagrams, and tables. The researcher suggests that additional research should be conducted on the impact of non-linguistic representations on item performance as similar results may support the redesign of assessments to amend the linguistic complexity and reduce the need for accommodation.

Another study (Wolf & Leon, 2009) utilizing DIF procedures on math and science items from 11 assessments for Grades 4, 5, 7 and 8 taken from 3 different states found results consistent to the previous study. The relationship exhibited between the proportion of language in an item and performance suggested that the impact of language was "somewhat lessened" when charts, visuals, or graphs were included in the items (pp. 156).

Ockey (2007) also utilized a DIF procedure on 10 items from the NAEP math test (also used by Abedi & Lord, 2001), and found only 1 item exhibited DIF against

ELLs. Based on his results, Ockey concludes that ELLs can be tested validly on large scale standardized math assessments (pp. 161). However, this interpretation should be approached with caution because the item set only included 10 items out of the original corpus of 69 items. The ten items used for Ockey's study were subjectively chosen by researchers because they seemed to be the most likely to confuse a student due to language complexities, however, students may be confused more by items that were not included.

Although these studies focused on the correlation between language complexity and student achievement on math tests, none of them dealt with the issue of defining a "word problem," as it pertained to their study. Most use the term to loosely describe any mathematics item that includes words as opposed to an item composed strictly of numbers and symbols. This is evidenced by the following statement in Shaftel et al. (2006). "Furthermore, all items were presented as word problems, though the number of words per item ranged from 2 words (in six items at 4th grade) to 177 words (in three items at 10th grade), with a mean of 45 words (pp. 110)." Most studies do not even provide a definition or description of what they consider to be a word problem. This lack of definition complicates interpretations because the cognitive and language processes required for solving what is thought of as a traditional algebraic word problem are much more complex than that of a straightforward computational problem described in a sentence.

Moreover, studies should continue to focus on the effect that non-linguistic representations have on student performance as they may be a suitable construct relevant accommodation. Thus the specific types of accommodations that are provided for ELLs to alleviate language impact are posed.

Accommodations used for ELLs to alleviate the language impact. Appropriate test accommodations for ELLs are intended to reduce the impact of language proficiency on content assessments such as math without changing the target of assessment (Kieffer et al., 2009) and that do not give them advantage over students who do not receive the accommodation (Abedi et al., 2004). The use of such accommodations could allow those involved with high stakes testing to make more valid inferences about an ELL's knowledge, however, these accommodations must be effective and valid. In regard to accommodations, Kieffer et al. (2009), describe effectiveness as "the extent to which students receiving the accommodation demonstrate improved test scores," and validity as "the notion that the accommodation should improve the performance of students who require it but not affect the performance of students who do not" (pp. 1171).

Since inclusion of ELLs on standardized tests occurred only recently, the empirical research base on which accommodations are valid and effective is minimal. Two comprehensive overviews (Abedi et.al, 2006; Rivera & Collum, 2004) and one meta-analysis (Kieffer et al., 2009) of the research to date on accommodations were reviewed. The studies to be discussed were narrowed down to those that specifically dealt with ELLs and mathematics tests. Research on bilingual dictionaries and glossaries were excluded because they only studied the accommodation with science tests.

The most commonly used accommodations for ELLs can be divided into two categories: modifications to the test and to the test procedure. The empirical research in these two categories will be discussed followed by an account of the current state policies on accommodating ELLs.

Modifications to the test and to the test procedure. The majority of the research on accommodations include some type of modification to the language of the test (Abedi, Lord, & Plummer, 1997; Abedi et al., 2006; Abedi et al., 2004; Johnson & Monroe, 2004; Kieffer et al., 2009; Rivera & Collum, 2004; Robinson, 2010; Sato, Rabinowitz, Gallagher, & Huang, 2010). The most researched modifications include linguistic simplification, English dictionaries or glossaries, tests in the native language and extended time.

Linguistic modification. Research has shown that mathematics test items can be linguistically modified to simplify the English language load without changing the construct being measured (Abedi, et al., 2006; Abedi et. al, 2004; Abedi & Lord, 2001; Johnson & Monroe, 2004; Sato, et. al, 2010), although results of effectiveness vary. According to Abedi and Lord (2001), shortening sentences, removing low frequency construct irrelevant vocabulary, replacing conditional clauses with separate sentences, and changing complex question phrases to simple question words are among some of the ways an item could be modified or simplified.

Abedi & Lord (2001) gave a 25 question paper and pencil test containing 10 original National Assessment of Educational Progress (NAEP) test items, 10 linguistically modified questions and five noncomplex control items to 1, 174 eighth grade students. Thirty-one percent of the students were designated as ELLs. Results showed that ELLs scored lower on the math test than proficient English speakers but that linguistic modification of test items resulted in statistically significant differences, albeit a small fraction (.17) of an item difference.

Using a similar method, Sato and colleagues (2010) gave one of two 25 question math tests (one linguistically modified and one with original state and nationally released items) to 4,617 middle school students divided into three sub-groups of English proficiency. Differences in effect of linguistic modification for the three sub-groups differed by the scoring approach used, with only one out the four approaches showing any significant differences. In another study, Johnson and Monroe (2004) found that simplifying the language on test items for ELLs did not make a difference in their performance with findings indicated that ELLs performed slightly better on original test items.

English dictionaries and glossaries. In the meta-analysis conducted by Kieffer et al. (2009), the use of English language dictionaries or glossaries was the only accommodation found to have a statistically significant and positive average effect size. However, the types and descriptions of dictionaries and glossaries varied by study leaving no identifiable standard to determine the use of a dictionary or glossary as an appropriate form of accommodation (Rivera & Collum, 2004).

Abedi, Courtney & Leon (2003) compared four different accommodations for math assessments given to 4th and 8th graders. One of which was a customized English dictionary with a paper and pencil test, while another utilized pop-up glosses on a computer test. ELLs performance on the computerized test was significantly higher than other accommodations, conversely, interpretation of the results must take into account the factor of using a computer for testing. Another study by Abedi, Hofstetter, Baker & Lord (2001) resulted in ELLs benefiting most from the combination of an English glossary of non-technical terms and extra time, although non-ELLs also benefitted from this combination. Of further note, the use of glossary only was the least beneficial for ELLs.

A study conducted by Wolf, Kim, Kao, and Rivera (2009) after the meta-analysis took place found no significant difference in the performance of ELLs on a math test with a glossary provided. Further insight into the ineffectiveness of the glossary was garnered through student interviews, which indicated that the students did not use the glossary for various reasons including forgetfulness and inexperience using the tool.

Tests in the native language. Accommodations in the native language include written translation of test directions and/or items, oral repetition of the directions and/or items in the native language using audiotape or dvd (Rivera & Collum, 2004). The majority of the studies to date have focused only on Spanish translation of assessments (e.g. Hofstetter, 2003; Robinson, 2010; Solano-Flores & Li, 2009).

In Robinson's (2010) study, ELLs in kindergarten and 1st grade performed significantly better on mathematics assessments in Spanish rather than English. Hofstetter (2003) found that students who took a Spanish version of the NAEP mathematics assessment generally scored slightly lower than those taking the standard test. Yet he notes that students getting math instruction in Spanish performed better on the Spanish version than students who received math instruction in Spanish but took the standard test.

One unique study of note was by Solano-Flores and Li (2009). The researchers went a step further and included language variation in the first language (L1) to study the accommodation of test translation on NAEP mathematics items for different groups of students belonging to the broad linguistic group of Spanish speakers. Math items were given to students in bilingual education classes in a city on either the U.S East Coast or the U.S. West Coast in English, standard Spanish, and local Spanish (Mexican or Dominican). The results showed that the main source of score variation was the interaction of student, item and language (pp. 192).

Extra Time. A minimal amount of studies have exclusively focused on modifications to test procedures for ELLs, although these are among the most commonly permitted accommodations on national and state tests (Abedi et al., 2003; Abedi et al., 2004;). The accommodation of extra time is the easiest and cheapest to administer and it does not require the revision of any test items or directions (Abedi et al., 2004), however, the research is inconclusive on whether the strategy is valid (pp. 12). Both ELLs and English proficient students benefited from extra time on an assessment of NAEP math items in a study by Abedi, Lord, Hofstetter, and Baker (2000). In the 2003 study by Abedi et al., 4^{th} grade ELLs performed better on a mathematics assessment with extra time over standard conditions, use of customized dictionaries, or small group administration.

On the whole, the accommodations studied, although valid, have shown inconsistent and minimal if any effects on reducing the performance gap between ELLs and non-ELLs. Since linguistic modification of construct irrelevant language has not led to improved performance, more research should be conducted on alleviating the effects of the construct relevant language of the mathematics register without changing the validity and reliability, such as the use of non-linguistic representations mentioned earlier. More research in this area is especially important as interpretation of scores with and without accommodation is further complicated by the inconsistent policies and use of accommodations for ELLs by state.

Current state policies for accommodating ELLs. In a study of state policies for ELL accommodations Rivera and Collum (2004) found that most states had used a taxonomy developed for students with disabilities (SWD) to organize the 75 accommodations listed for ELLs. Forty-four of those accommodations were responsive to the needs of ELLs, while 31 were relevant only for SWDs (pp. 19). Including ELLs in the taxonomy for SWDs could obscure the differences in appropriateness of the accommodation for either sub-group of students. Furthermore, the policies vary in number, type, and use for certain content areas by state.

The State of Michigan currently offers a total of 63 accommodations organized in the taxonomy for SWDs (Office of Educational Assessment and Accountability [OEAA], 2009). Of the 63 accommodations, 23 are considered standard for ELLs and 9 of those are coded as *universal,* meaning any student is allowed to use those accommodations (i.e. use of a ruler supplied by the state, use of a highlighter).

Apart from the use of a word-for-word bilingual glossary, any accommodations providing direct linguistic support (i.e. directions or problems read aloud by test administrator) can only be used for ELLs that are: dominant in their native language and at a basic or lower intermediate level of English proficiency. In order to receive any direct linguistic support in the native language, the student must receive bilingual services at school to maintain their native language and be at a basic or lower intermediate level of English proficiency. There are no complete tests written in any languages other than English and there are no simplified English versions of the tests for ELLs.

Summary

Recognizing and understanding the relationship between language and mathematics learning for ELLs is crucial for educators and anyone associated with large scale assessment design and score interpretation. Incorporated with a socio-cultural perspective, the descriptions of the mathematics discourse by Lemke (2003) and O'Halloran (2000) constitute what is referred to in this study as a mathematics register. Learners must utilize the resources of mathematical symbolism, visual display, and natural language interdependently to negotiate mathematical meaning.

ELLs have the challenge of translating between three registers (native language, English, mathematics) to successfully solve math problems. This process includes decoding natural language, content-specific terminology (vocabulary and syntax), and non-linguistic representations. Furthermore, Brown, Cady, and Taylor (2009) identified several instances in the U.S. mathematics register that may cause confusion for ELLs including: the use of a decimal comma instead of a decimal point to indicate place value, the way that dates are written, and the use of customary units as opposed to metric units.

Another important observation concerning the challenges ELLs face with the mathematics register is that students in the study conducted by Lager (2006) had a sense of “false knowing” in regard to vocabulary in math problems. In other words, the students thought they understood more of the language than they actually did as indicated by their performance.

Research has also focused on the impact that construct-irrelevant language features of math items from high stakes tests has on ELL performance. The majority of these studies have found a relationship between language complexity and performance on math test items, but item length was the only language feature to consistently show a negative effect on performance (Martiniello, 2009).

Two studies conducted by Martiniello (2008, 2009) analyzed 10 items from the MCAS that showed DIF disfavoring ELLs. Further examination of these items confirmed that linguistic complexity (i.e. polysemous words, multiple clauses) could negatively influence the performance of an ELL. The researcher also found that half of the items showing DIF were from the learning strand Data Analysis, Statistics, and Probability. Results from the second study showed that the impact of linguistic complexity was weakened with the inclusion of a non-linguistic schematic representation. Martiniello suggested that further research be conducted on the relationship between item performance and both the Data Analysis, Statistics and Probability strand and non-linguistic representations.

Appropriate accommodations are intended to reduce the impact of construct irrelevant language features without disadvantaging students that do not receive those accommodations and without changing the target of assessment (Kieffer et al., 2009). One of the most studied accommodations is that of linguistic modification, or simplified English (Abedi, et al., 2006; Abedi et. al, 2004; Abedi & Lord, 2001;

Johnson & Monroe, 2004; Sato, et. al, 2010), but the overall outcome is inconclusive due to varying results of effectiveness. The meta-analysis conducted by Kieffer et al. (2009) determined that the use of English dictionaries or glossaries was the only accommodation studied to have a statistically significant and positive average affect size. Since the research regarding ELLs and high stakes testing is fairly recent, state policies on accommodation vary greatly in number, type and use for certain content areas (Rivera & Collum, 2004).

Conclusions

Valid inferences about ELLs performance on mathematics high stakes test are critical due to the consequences attached for ELLs and schools. The available research is minimal and for the most part inconclusive, perhaps because more of the issue is with the complexity of the construct-relevant language features. Further research into the effect of nonlinguistic representations on language impact is critical, as increased inclusion of such could act as a construct-relevant accommodation. Additionally, a comprehensive exploration of the relationship between item strand and performance is necessary. Any strand that consistently has a negative effect on performance should be examined for its accessibility to all students.

Since ELLs must acquire both the academic English register and the mathematics register simultaneously, teachers must be prepared to act as both a language and content area teacher to scaffold learning in the zone of proximal development. This has great implications for teacher preparation schools, professional development creators, and administrators. Finally, a consistent definition or description of what constitutes a ‘word problem’ must be provided and agreed upon by experts in the field in order for research results to be interpreted validly.

Chapter Three: Research Design

Introduction

The aim of this study is to investigate the extent to which language proficiency influences English language learners' performance on mathematics high stakes tests as compared to their proficient English speaking peers. The central research question of the study is: To what extent does language proficiency influence achievement on high stakes mathematics tests? This broad question is broken down into the following four specific sub-questions:

5. Do English language learners perform differently on computation questions as opposed to word problems on high stakes mathematics tests? How does their achievement compare to non-English language learners on these tests?

6. Is the difficulty of an item, due to language complexity, alleviated with the inclusion of graphical, pictorial, or schematic representations?

7. Are there specific strands, as defined by the Michigan Curriculum Framework (Data & Probability, Algebra, Number & Operations, Geometry, Measurement), that English language learners perform differently on in comparison to non-English language learners on high stakes test items?

8. What language issues do students identify as confusing or unfamiliar in high stakes math test items?

In order to answer each question, the study included statistical analyses of high stakes test item performance data along with an analysis of questionnaire responses to provide triangulation. This chapter will begin with the design and rationale of the study and followed by descriptions of the archival data, participants, instrumentation, data collection, and data analysis procedures. A short summary of the research design will conclude the chapter.

Design and Rationale of the Study

This study utilized a cross-sectional study design and implemented a quantitative research method to investigate the issue. This design is used to study data collected all at once from an entire population or subset of that population. A cross sectional design was chosen because one can collect large amounts of data quickly and cost effectively, and use it to investigate relationships between several variables among different groups naturally found in the population (Gay, Mills & Airasian, 2009).

Archival Data and Participants

A description of the population, samples and sampling criteria for both the archival data and the questionnaire research, and the role of the researcher are provided in this section.

General population. The general population for this study consists of 7th and 8th grade students in a county intermediate school district (ISD) in West Michigan. No studies have been conducted to examine the impact of language on performance on high stakes mathematics tests in Michigan, which is necessary due to recent surges in the ELL population and consequences for poor performance on high stakes tests. This county ISD was chosen because it had a diverse population of 7th and 8th grade students including a significant amount of ELLs (6%) as compared to the state average of 3%. Gender was represented evenly with 51% males and 49% females. Data on ethnicity showed that about 68% of the population were White, 15% Black, 12% Hispanic, 3% Asian/Pacific Islander, and 2% were Multiracial.

Archival data set. The data of this study were obtained from a county intermediate school district (ISD) in West Michigan for the 7th and 8th grade 2007, 2008, and 2009 Michigan Educational Assessment Program (MEAP) tests. In order to study the difference in performance on mathematics items between ELLs and non-ELLs, purposeful sampling procedures were applied to select the data set for analysis. The method was chosen for convenience, time and cost efficiency, and because a large data set could be obtained. The data sample represented the tests taken by 24,693 seventh and eighth grade students, from 17 out of 20 public school districts in the county ISD, who took the MEAP test in 2007, 2008, or 2009. Three school districts were excluded from the sample because they did not have any ELLs take the MEAP during those years.

The archival data set categorized performance by item for two groups: ELL and non-ELL. Table 1 presents the size of each group comprising the sample of test takers. Former ELLs were included in the non-ELL category as there was no way to identify this sub-group in the data given to the researcher. As expected, the data points from non-ELLs greatly outnumber that of ELLs; however, since so many data points were provided and the demographic is similar to that of the population the bias is reduced. Furthermore, the demographics of each of the school districts varied greatly so the entire data set was used so that each sub-group would be represented making the results more generalizable to other ISDs with similar demographics in the state of Michigan.

Table 1
Breakdown of MEAP Test Takers by Year and English Language Proficiency Status

Year	Non-ELL	ELL	Total
2007	11,763	822	12,585
2008	11,286	822	12,108
Total	23,049 (93%)	1,644 (7%)	24,693

The items that operated as the unit of analysis were released items from the 2007 and 2008 seventh and eighth grade MEAP tests. Only 148 of the original 263 items from the four MEAP assessments were released publicly. Forty-four of the publicly released items assessed future or extended core curriculum and were not given to all students, so only the remaining 104 released items were included in the study. All of the items were in multiple choice format.

The criterion-referenced MEAP assessments are based on the Michigan Grade Level Content Expectations (GLCEs) and are designed to measure what Michigan educators believe all students should know and be able to achieve in the content areas (MDE: Office of School Improvement, 2006). The GLCEs for mathematics are divided into five main strands: Number and Operations, Algebra, Measurement, Geometry, and Data and Probability.

According to the 2008-2009 MEAP Technical Report (MDE: Office of Assessment and Accountability, Pearson Educational Measurement, and Harcourt Assessments, Inc, n.d.), empirical item response theory (IRT) reliabilities for the seventh and eighth grade math MEAP range from 0.85-0.89 and 0.86-0.88, respectively (p. 138). The seventh and eighth grade MEAP math assessments were reviewed by 6 or 7 mathematics experts to analyze agreement between the GLCEs and assessment items to provide a measure of content validity (p. 147). Two DIF analyses, Editorial Bias Review and Mantel-Haenszel Delta tests, were performed by experts on each item to identify any items on which members of a focal group have a different probability of answering an item correctly from members of a reference group after being matched by means of ability level. The statistical DIF analysis conducted on the MEAP only compared males versus females, and white versus black (p. 153).

Questionnaire participants. The accessible population for the questionnaire included all of the seventh and eighth grade students (N=224) from the middle school at which the researcher was employed. The school is located in the county ISD from which the archival MEAP data was obtained. Permission to include students in the study was granted by the principal (see Appendix A). The gender distribution was balanced with 52% females (N = 117) and 48% males (N=107). Forty-one (18%) of the students were ELLs, and 195 (87%) of the students were disadvantaged or of low socio-economic status.

The sample was selected through non-proportional quota sampling procedures to include students with varying gender, mathematics grade, ethnicity, socio-economic, and linguistic backgrounds. The 7th and 8th grade mathematics teachers were the primary informants of the demographic distribution of students in each of their classes and provided students with number codes so that their identity would be protected. The sample comprised of 61 students from one 7th grade class (N = 30) and one 8th grade class (N = 31).

Out of the 61 students, 30(49%) were male and 31 (51%) were female. Nine (15%) of the students were ELLs, all with Spanish as their primary language. Thirteen (21%) of the students were White, 34 (56%) were Hispanic, 7 (11.5%) were Black, and 7 (11.5%) were Mixed Race.

The role of the researcher. The researcher in this study served as a document analyst and questionnaire administrator. As a document analyst the researcher was an objective evaluator and no actions was undertaken to influence the data, which could not have been modified given that the tests used in the study had been administered 3 to 4 years earlier. In addition to being an archival analyst, the researcher also performed the role of test administrator and abstained from being involved with students during the test taking period. Since the researcher worked at the middle school and was familiar to the participants, any distortions produced by the presence of the researcher as a questionnaire administrator were minimal to none.

Instrumentation

The instruments used in the study included set of coded items from four middle school MEAP mathematics assessments from 2007 and 2008, and a questionnaire developed by the researcher.

MEAP mathematics test items. To investigate the difference in performance by ELLs and non-ELLs on different mathematics test items, a protocol was developed to code items by type before analysis. The protocol directed raters to code each of the released MEAP items used in the data set using the following definitions.

- Word (W) - a word problem is considered a mathematical question presented in words and graphics which must be translated into mathematical and/or algebraic symbols in order to correctly determine the solution to the question.

- Computation (C) - a computation problem is considered a mathematical question or statement presented in words and graphics which can be solved without translating the words into mathematical or algebraic symbols. (i.e. *Solve for x, Evaluate, Determine the product of* 4×3). This definition is expanded to include definition and identification problems (i.e. *Which of the following is a square? What is the coordinate of D)* which are neither word problems nor computation problems in order to distinguish a dichotomous variable: word problem and non-word problems.

- Graph (G) – any item coded as G contains a non-linguistic representation (i.e. table, graph, picture, diagram, geometric figure).

- No Graph (N) – any item that consists solely of words, numbers, and mathematical notation and symbols.

The items were coded independently by the researcher and two math experts as either word (W) or computation (C) and as either including a non-linguistic representation (G) or not (N). The codes were compared between the raters for any discrepancies. Only two items were found to have disagreement and consensus for recoding was determined after discussion. Table 2 presents the breakdown of the items by year, grade, and code as determined by the raters.

Table 2
Breakdown of Released MEAP Items by Year, Grade, and Code

Item Type	2007		2008		Total
	7th grade	8th grade	7th grade	8th grade	
Computation	13	7	13	8	41
Computation with Graph	5	2	1	2	10
Word	11	9	9	7	36
Word with Graph	1	11	2	3	17
Total	30	29	25	20	104

The original data set included information that was previously disaggregated into item level performance by group (ELL or non-ELL) and organized by test cycle, district, grade level, item number and GLCE. Table 3 shows an example of how the data was organized. The data set was stored electronically so that it could be uploaded into SAS 9.2 for statistical analysis.

Table 3
Example of MEAP Mathematics Item Performance Data Set from the ISD

Test Cycle	District Code	Grade Code	LEP[1] Status	Item Number	GLCE	Student Count	Percent Passing
Fall 2007	41	7	LEP	49	A.FO.06.12	45	11.11%
Fall 2007	41	7	Not LEP	49	A.FO.06.12	97	35.05%

Note. The district codes were changed so that the school districts could not be identified.
[1] The data set refers to ELLs as Limited English Proficient (LEP) and non-ELLs as Not LEP.

Questionnaire. The questionnaire was developed by the researcher specifically for this study. A description of the development, translation and piloting procedures are provided in this section.

Development. The researcher designed a questionnaire which included a similar procedure to one used by Lager (2006), in which participants were asked to answer each of the items and to circle any confusing or unknown words or phrases. The researcher adapted this procedure to include unknown symbols and non-linguistic features. The questionnaire consisted of three sections (see Appendix B).

The first section included five word problems indentified by the Grade Level Content Expectation (GLCE) Strand from the 2007 and 2008 seventh and eighth grade publicly released MEAP test items. Two of the items were from the Data and Probability Strand, one of which included a non-linguistic feature. The remaining three items were from the Number and Operations, Geometry, and Algebra strands.

The Geometry item included the other non-linguistic feature. The Measurement strand was not included because there are no GLCEs for this strand in the 7th and 8th grades at the present time.

The second section consisted of three open-ended questions regarding the math items in the previous section and one Likert-scale type question regarding pictures, graphs and tables. The third section requested demographic information including language status, gender, age, math grade, home language/s, and the grades in which the participant attended school in the U.S.

Translation. The directions along with the second and third sections were translated into Spanish by a bilingual Spanish-English expert and then verified by two other bilingual Spanish-English experts. The math items in the first section were not translated into Spanish because all students in the state of Michigan, regardless of language status or length of time in the country, must take the MEAP mathematics assessment written in English, which is an issue of focus in this study. Bilingual glossaries were available for all ELL students because they are an approved standard accommodation for MEAP assessments (OEAA, 2009).

Pilot. In order to collect and analyze emerging issues or patterns, the questionnaire was piloted with ten 7th and 8th grade students from a nearby school varying in linguistic background, socio-economic status and gender. The pilot sample participants consisted of six 7th grade students (2 males, 4 females), and four 8th grade students (2 males, 2 females). Three of the students were ELLs (two native Spanish speakers, one native Korean speaker), six students were former ELLs (4 native Vietnamese speakers, 2 native Spanish speakers and 1 native Bosnian speaker) and one student was a native English speaker. All students had the option of taking the questionnaire in English or in Spanish, and all ten students chose to take the questionnaire in English. After the questionnaire was administered, the students were asked for feedback on the clarity of directions and questions. The results and feedback were reviewed by the researcher and a math teacher at the school and three modifications to the questions were deemed necessary.

Students were unsure how to answer the second part (in bold print) of the following two questions from the second section:

1. Which problem was the easiest for you? **Please explain why.**
2. Which problem was the most difficult for you? **Please explain why.**

These two questions were changed to the following:

1. Which problem was the easiest for you? **What made it easier than the others**?
2. Which problem was the most difficult for you? **What made it more difficult than the others?**

The Likert-scale question was eliminated because it was the only one of its type, seemed biased, and the majority of the students chose "neutral" as the response. The final version was divided into the same three sections with changes made only to the second section. After the modifications were made, translation into Spanish was completed using the same procedure as before.

Data Collection

The archival and questionnaire data collection occurred separately on two different occasions. The process for gathering the data is described in this section.

Archival data set. The 2007, 2008, and 2009 MEAP mathematics item performance data set for 7^{th} and 8^{th} grade students was obtained with permission from the data consultant at a county ISD in West Michigan (See Appendices). The original data set included information from 20 school public school districts and was sent electronically to the researcher in spreadsheet format.

Questionnaire response collection. Prior to the administration of the questionnaire to the participants, permission from the Human Research Review Committee was obtained (see Appendix C). Parental or legal guardian permission also had to be obtained in order for any student under 18 to participate in the study (see Appendix D). Parental consent forms in English and Spanish were distributed to the 61 students selected for the sample after the researcher explained the objectives and purposes of the study to both classes**.** The consent form was translated from English into Spanish using the same procedure as the questionnaire translation. At the end of the consent form, the student was asked to check one of two boxes indicating whether they wanted to participate in the study or not and were asked to sign the form in either instance. In addition, if the student opted to participate they were asked to indicate whether they wanted the questionnaire in Spanish or English. The form also had to be signed by the student's legal guardian or parent, authorizing permission for their child to participate.

The classes were informed that the questionnaire would take about 15 minutes and would be administered during their math class two days later. Anyone that did not choose to participate or did not obtain parental consent would be given a separate math assignment by their teacher.

After two days, 18 out of the 61 students in the sample returned with signed consent forms, two of which selected not to participate. Since only 16 students selected to participate, any findings must be generalized with caution. One explanation for the small return size was the short amount of time available to send home the consent forms and administer the questionnaire. The master's thesis is to be completed in one semester and much of that time was spent in the IRB process. The time frame for questionnaire administration could not be extended any longer because the school was closed for spring break vacation soon after the questionnaire was administered. As a result, the researcher continued the questionnaire with the small but acceptable sample size (27% of the students invited to participate).

Nine of the participants were from 7th grade: 3 males and 6 females. The remaining 7 participants were from 8th grade: 6 females and 1 male. Six of the participants were ELLs, one was classified as advanced, three as high-intermediate, one low-intermediate, and one that was not sure of their level of English proficiency. All participants have attended school in the U.S. since kindergarten. Languages spoken at home included English and/or Spanish including one participant that spoke Portuguese in addition to Spanish and English, and one that spoke French in addition to Spanish and English. Only 1 out of the 16 participants requested that the questionnaire be in Spanish.

The questionnaire was distributed to the participants in their math classroom. Each student had a calculator and access to a bilingual glossary. Verbal instructions were given by the researcher and a trained Spanish-English bilingual paraprofessional. Participants were given as much time as needed to complete the questionnaire, although many finished within the estimated 15 minute time-frame. The questionnaires were collected by the teachers and the researcher and kept in a file folder in a locked cabinet until data analysis could be conducted.

Data Analysis

This study analyzed item performance data by group from four different MEAP assessments using a logistic regression model to determine if the type of item differentially affected the rate that ELLs and non-ELLs answered an item correctly. Questionnaires, including items from the mathematics MEAP assessments, were administered in order to determine any language features of math items identified as problematic by students rather than researchers. Descriptions of the analysis for both data sets are provided in this section.

Logistic regression tests. The original data set was read using SAS 9.2 and then trimmed to contain only the data for the 104 released items from the 2007 and 2008 test cycles. All of the data for the 2009 test cycle was eliminated because the assessment items were not released for that year. The information from three districts was also eliminated because they did not have any item performance information for ELLs. The resulting data set contained 3,386 observations.

Before conducting the analysis the new variables, as a result of the coding protocol, were entered and multiple data checks were made to ensure that the coding of new variables in SAS 9.2 was done properly.

The appropriate test for this study was a logistic regression model because the response variable was dichotomous and there were multiple predictors. If there were only one or two explanatory variables then a Chi-square test of independence or a Cochran Mantel-Haenszel test could be used, respectively. However, since there are more than two explanatory variables along with their interaction terms, then logistic regression is the analysis capable of capturing this design.

An important caveat to the analysis is that each of the 3,386 observations was treated independently even though in reality this was not the case because the same students are answering multiple questions. This was a necessity due to the nature of the data provided, which did not give information at a student level; thereby, making it impossible to measure the within-subject effect. Although the amount of data points remain the same, the within-subject effect of each student is not factored into the model which violates an assumption of the design. Again, this could not be avoided because of the way the data was provided.

The response variable is whether the student passed an item and the explanatory variables and their codes are as follows:

- English language proficiency status: ELL or non-ELL (LEP[1])
- item type: word (W) or computation (C)
- inclusion of non-linguistic representation: graphic (G) or no graphic
- the interaction effect between LEP and CW
- the interaction effect between LEP and G

The interaction effects are the main interest in the study because these will answer the first and third research questions whether ELL students perform differently on computation and word problems compared to non-ELL students and likewise with problems including non-linguistic representations or not.

A second analysis using the same logistic model was performed to see if the GLCE strand of the item had an effect on how ELLs and non-ELLs perform using the strand as a predictor. In addition, a third analysis was performed with the GLCE

[1] LEP is used here because that is term used in the data from the ISD

strand variable dichotomized into a new variable of whether the item was a Data and Probability question or not. These analyses will answer the second research question.

Questionnaire response analysis. The initial analysis evaluated the math items for correctness and cross tabulated to study interaction effects on performance between ELLs and non-ELLs and performance by strand. Additionally, a frequency analysis was conducted on the words and phrases circled by students in order to answer the fourth research question. The answers to the three open ended questions in the second section were coded into one of three categories (easy, difficult, and strategy) by the researcher and analyzed in order to further investigate reasons for misunderstandings and strategies for dealing with those misunderstandings.

Although the sample is unintentionally small and gender biased, the results provide more insight on how the mathematical language on the MEAP items is interpreted by students. Another limitation is that students may have previous experience with some of the test items as these are often used in test preparation each year.

Timeline

The following is an outline of the research study timeline:

- Permission to conduct research with students was obtained in January 2011 from the middle school principal.
- Initial contact with the West Michgian ISD data consultants was made in the last week of January, 2011. The data set of MEAP item performance for 7th and 8th grade students in the county was provided on January 27, 2011 along with permission to use the data set for research.

- Research proposal was sent to the IRB board for approval on January 25, 2011.
- Questionnaire was developed, piloted, revised and translated between January and March, 2011
- Item performance data analysis was conducted at the Statistics Consulting Center in March of 2011.
- Research proposal approved by IRB board on March 24, 2011.
- Consent forms were distributed and questionnaires were administered to students with signed parental consent forms in late March, 2011.
- Questionnaire responses were tabulated, coded and organized, and analyses were conducted during the first week of April.

Summary

This study utilized a cross-sectional research design. The issue was investigated using quantitative research methods along with analyses of questionnaire responses to triangulate findings and obtain a more complete picture of the issues being studied. A logistic regression model was implemented on a data set of group performance by item on MEAP mathematics assessment from 17 school districts in a Michigan ISD. Interaction effects between ELLs and item type, item language, and item strand were the main interest of the statistical analysis. Questionnaires related to the math items in the statistical analysis were administered to 16 participants after IRB and parental permissions were obtained. Results were studied by cross tabulation, content and frequency analyses and later compared to the results of the statistical analysis.

Chapter Four: Results

This chapter presents the findings from the study. First the context of the study including the demographics of the participants is discussed followed by the findings which are presented in two sections. A short summary is provided to conclude the chapter.

Context

This study collected and analyzed data from two different sources: archival data from the 2007 and 2008 MEAP assessments, and a questionnaire. The demographic characteristics for each source are described below.

Characteristics of the 2007 and 2008 MEAP test takers. The data set provided to the researcher included item performance scores from 20 school districts for 7th and 8th grade students for the 2007, 2008, and 2009 MEAP assessment. Data from 3 of the 20 school districts were eliminated because they had no performance data for ELLs. All data from 2009 was also eliminated because there were no publicly released MEAP items for that year.

Table 4
Year and Grade Level Breakdown of 2007 and 2008 MEAP Test Takers

Year	*7th grade*		*8th grade*		*Total*
	Non-ELL	*ELL*	*Non-ELL*	*ELL*	
2007	5,870	429	5,893	393	12,585
2008	5,596	412	5,690	410	12,108
Total	11,466	841	11,583	803	24,693

The data set only identified students by grade level (7^{th} or 8^{th}) and as LEP[2] or Not-LEP. Other demographic information including level of English proficiency, although of interest, was not available. Table 4 shows a breakdown of the test takers by grade level, year and English proficiency status. One of the school districts included in the archival data set is the setting where the questionnaire was administered.

Characteristics of the Questionnaire Participants. Sixty-one students were invited to participate in the questionnaire. Eighteen students returned with signed parent consent forms. Two selected not to participate bringing the total participant count to sixteen. Fifteen questionnaires were completed and one survey was complete except for one response. The same survey also had responses that did not appropriately answer the remaining two short answer questions. Nevertheless, all data collected will be considered, as the blank response and inappropriate responses were on short answer questions that do not affect the results of the analysis. The particular participant may not have understood the questions even though they were not an ELL.

Table 5 provides gender and English proficiency status ratios by grade level. Six of the participants were ELLs, one was classified as advanced, three as high-intermediate, one low-intermediate, and one that was not sure of their level of English proficiency. Nine (56%) of the participants use both Spanish and English at home, one of which also spoke Portuguese and another also spoke French. Two (13%) participants use Spanish only as their home language and 5 (31%) participants use English only as their home language.

---

[2] The data set refers to ELLs as Limited English Proficient (LEP) and non-ELLs as Not LEP

Overall, 82% of the sample was eligible for free and reduced lunch. Socio-economic status of each individual participant could not be included since that information was purposefully not requested on the questionnaire. Students may not know their socio-economic status or may be embarrassed to state it even though the questionnaire was anonymous.

Table 5

Breakdown of Questionnaire Participants by Grade Level, Gender, and English Proficiency

	Female	Male	ELL
7th	6	3	4 (2 *H.I,*1 *L.I*, 1 *A*)
8th	6	1	2 (1 *H.I.*, 1 *N.S*)

Note. H.I. = High Intermediate; *L.I.* = Low Intermdiate; *A* = Advanced; *N.S.* = Not Sure

The characteristics and demographics of the participants are necessary to generalize the findings.

Findings

The findings of this study are presented in two sections. The first section addresses the first three research questions and shows how the type of problem differentially affects the rate that ELLs and non-ELLs answered an item on the MEAP mathematics assessment correctly through logistic regression procedures. The second section adds more understanding to the first three questions and addresses the fourth question of which language features of test items are identified as problematic by questionnaire participants.

Logistic regression analyses. Findings regarding the first two sub-questions will be discussed first since the results overlap. Those questions are as follows:

1. Do English language learners perform differently on computation questions as opposed to word problems on high stakes mathematics tests? How does their achievement compare to non-English language learners on these tests?

2. Is the difficulty of an item, due to language complexity, alleviated with the inclusion of graphical, pictorial, or schematic representations?

Results from the logistic regression analysis performed using SAS 9.2 show that there is a significant difference in passing rates between ELLs and non-ELLs depending on the item type and the presence of non-linguistic features. Table 6 shows that all three main effects of language proficiency status, item type, and inclusion of non-linguistic feature are very significant with p-value <.0001 (α = .05). In other words, ELLs and non-ELLs have different passing rates on the MEAP mathematics assessment and all students have different passing rates on computation and word items, as well as items including or excluding non-linguistic features.

Table 6

Analysis of Maximum Likelihood Estimates for Language Proficiency Status, Item Type, and Inclusion of Non-linguistic Features and Their Interactions

Parameter			*df*	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept			1	0.00398	0.00581	0.4678	0.4940
LEP/EP	EP		1	0.3529	0.00581	3684.2942	<.0001
Comp/Word	C		1	-0.0282	0.00516	29.9161	<.0001
Graph/No Graph	G		1	0.0698	0.00587	141.7570	<.0001
lep*cw	EP	C	1	-0.0312	0.00516	36.5114	<.0001
lep*graph	EP	G	1	-0.0233	0.00587	15.7904	<.0001

Note: LEP and EP are used here instead of ELL and Non-ELL because that is term used in the data from the ISD.C= computation item; G = non-linguistic features; lep*cw = language proficiency status with computation or word item; lep*graph = language proficiency status with non-linguistic feature

Moreover, the interaction effects of English proficiency status with item type and with inclusion (or exclusion) of non-linguistic features are also very significant with p-value <.0001. Specifically, the difference in passing rates changes between ELLs and non-ELLs based on the item type and inclusion of a non-linguistic feature.

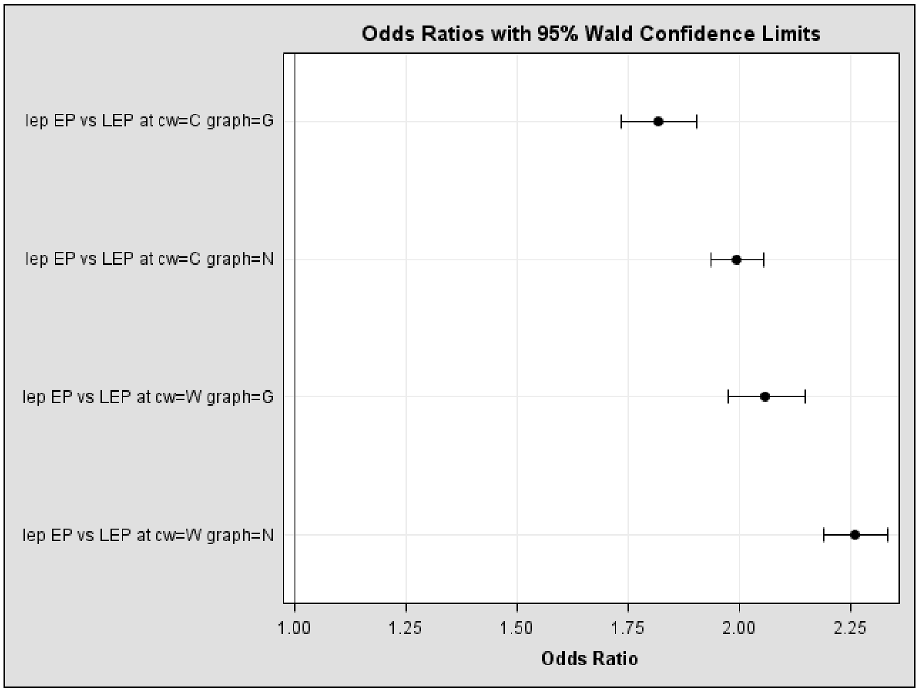
Table 7 and Figure 1 further illustrate the differences by providing the odds ratios of the groups. Each odds ratio represents the magnitude of how much greater the odds of passing are for non-ELLs than for ELLs in each of the following four cases: 1) computation problem with a non-linguistic feature, 2) computation problem without a non-linguistic feature, 3) word problem with a non-linguistic feature, and 4) word problem without a non-linguistic features. As can be seen more clearly in Figure 1, the performance gap between ELL and non-ELL students is greatest on word items without non-linguistic features. The gap is closest on computation problems that include a non-linguistic feature. Note that none of the confidence intervals (CI) overlap, indicating that they are all significantly different from one another.

Table 7

Wald Confidence Interval for Odds Ratios for Four Interaction Effect Cases

Label	Estimate	95% Confidence Limits	
lep EP vs LEP at cw=C graph=G	1.816	1.734	1.903
lep EP vs LEP at cw=C graph=N	1.994	1.936	2.054
lep EP vs LEP at cw=W graph=G	2.058	1.973	2.146
lep EP vs LEP at cw=W graph=N	2.259	2.187	2.332

Figure 1. Plot of Odds Ratios Showing Performance Gap between ELLs and non-ELLs for Four Cases



A second analysis using the logistic regression model included the GLCE Strand as another predictor. These findings answer the third research sub-question:

3. Are there specific strands, as defined by the Michigan Curriculum Framework (Data & Probability, Algebra, Number & Operations, Geometry, Measurement), that English language learners perform differently on in comparison to non-English language learners on high stakes test items?

The results showed that the different GLCE Strands are significantly different from one another with p-values <.0001. However, the interaction terms of the GLCE Strand with ELLs and non-ELLs are very insignificant (p-values ranging from .2424 to .7821). This indicates that the difference in passing rates between ELLs and non-ELLs is not affected by the GLCE Strand the item is based upon.

Table 8
Analysis of Maximum Likelihood Estimates for the GLCE Strand of Data and Probability and Interaction Effects

Parameter			*df*	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept			1	-0.1060	0.00894	140.4984	<.0001
LEP/EP	EP		1	0.3497	0.00894	1529.1095	<.0001
Comp/Word	C		1	-0.0577	0.00544	112.4159	<.0001
Graph/No graph	G		1	0.0883	0.00597	218.7031	<.0001
Data & Prob/Not	0		1	0.1507	0.00916	270.5928	<.0001
lep*dp	EP	0	1	0.00584	0.00916	0.4059	0.5241
lep*cw	EP	C	1	-0.0328	0.00544	36.3751	<.0001
lep*graph	EP	G	1	-0.0219	0.00597	13.4316	0.0002

In order to determine whether the GLCE Strand of Data and Probability had an impact on the likelihood of answering an item correctly for ELLs or Non-ELLs, a third analysis was conducted by dichotomizing the GLCE Strand into a binary variable: Data and Probability, and Not-Data and Probability. As shown in Table 8, the main effect of whether the item was from the Data and Probability Strand or not is significant with a p-value <.0001, denoting that all students perform differently on that type of question. On the other hand, the interaction term is insignificant at p-value = .5241. Therefore, the difference in passing rates between ELLs and non-ELLs in not significantly affected by whether an item is from the Data and Probability Strand or not.

Questionnaire response analyses. A discussion of the descriptive statistics, the confusing and unfamiliar words identified by participants, and the strategies they employ are included in this section.

Descriptive statistics. Table 9 gives a summary of total percent of correct responses stratified by English proficiency status. As can be seen, there is a gap between the performance of non-ELLs and ELLs. Out of a possible score of 5, the average raw score for ELLs was 2.35 and for non-ELLs was 3.4. The standard deviations indicate that scores for non-ELLs are closer together whereas there is more variability in the ELLs' scores.

The participant math grades from the previous trimester are strongly correlated (r =.659, p = .05) to their percent score on the five test items included in the questionnaire. The school is on a trimester schedule, in which students receive grades 3 times a year as opposed to 4. The grading scale ranges from an A+ (100%) to an E (59%-0%). The average math grade for the previous trimester was a B- , with a range from E to A. The math letter grade breakdown is as follows: three A's, one A-, four B's, two B-'s, one C+, two C's, two D's, and one E.

English language learners were outperformed by the non-ELLs on four of the five questions (see Table 10). Those four questions were based the Geometry, Number and Operations, and Data and Probability Strands. The question on which ELLs outperformed non-ELLs was based on the Algebra Strand. The highest group performance (85%) by ELLs was on the Geometry item with a non-linguistic feature, yet the lowest ELL group performance was on a Data and Probability question with a non-linguistic feature.

Table 9
Total Number of Students With Percent Correct According to English Proficiency Status

Percent Correct	0%	20%	40%	60%	80%	100%	Grand Total	*M*	*SD*
Non-ELLs	0	0	2	4	2	2	10	0.68	0.219
ELLs	1	1	0	3	1	0	6	0.47	0.305
Grand Total	1	1	2	7	3	2	16	0.60	0.250

Confusing words or phrases. Participants were asked to circle any confusing words, phrases, or pictures in order to answer the fourth research sub-question :What language issues do students identify as confusing or unfamiliar in high stakes math test items?

Nine out of the sixteen students circled words, phrases, or symbols that were unfamiliar or confusing, 8 of which were non-ELLs while only one was an ELL. Participants may not have circled confusing or unfamiliar words or phrases for several different reasons such as the following: unclear directions, misunderstanding, embarrassment, or inexperience with the procedure. This issue could be further explored through follow-up interviews and repeated cross-sectional studies. Table 10 shows the frequency of the words, phrases, and symbols circled by students per item, percent correct, strand, and inclusion of non-linguistic features. Analysis for each item will be discussed individually. The definitions provided by Wolf and Leon (2009) were used to differentiate between two sub-groups of mathematics specific academic vocabulary. The researchers used the terms *context-specific* and *technical* for the different categories of vocabulary. Context-specific vocabulary consists of words that are used in everyday life, but have a specific meaning when used in a discipline and are highly context dependent (i.e. face, square). Technical vocabulary consists of words that are specific to one discipline and are rarely used outside of that context (i.e. denominator, apothem).

Table 9

Frequency Chart of Confusing or Unfamiliar Language Features Circled by Participants and Compared by English Proficiency Status, Strand and Inclusion of Non-linguistic Features

Circled Language Features	Frequency	Number and percent correct by group	Strand	Non-linguistic feature
Item One:				
segment symbol in $\overline{LM}$	1	ELL - 5 (85%)	Geometry	Yes
congruent	1	non-ELL – 10 (100%)		
corresponds	1			
Item Two:				
six-sided number cube	1	ELL – 3 (50%)	Data and	No
numeral	4	non-ELL – 9 (90%)	Probability	
numerals	1			
Item Three:				
relative frequency	4	ELL – 1 (17%)	Data and	Yes
chose	1	non-ELL – 6 (60%)	Probability	
Item Four:				
Celsius	2	ELL – 4 (67%)	Algebra	No
represents the Celsius	1	non-ELL – 4 (40%)		
-13ºC, 6 ºC	1			
Item Five:		ELL – 1 (17%)	Number	
25, 2¼, 12½	1	non-ELL – 5 (50%)	and	No
			Operations	
Total	19			

Item One. Only one participant (an ELL) did not answer number one correctly and a majority of the participants (63%, N = 10) chose this item as the easiest (see Tables 11 and 12 for the easiest and most difficult items chosen by group);. A sample (see Appendix E for a complete listing of comments) of the reasons participants gave for choosing number one as the easiest item follow:

Comment 2 (of 16): *Number One was the easiest because I really understand triangles and the others were more difficult. (non-ELL)*
Comment 6 (of 16): *Number one was the easiest. Because you just had to flip it.(ELL – High Intermediate)*
Comment 11(of 16): *Number 1, I could see it and I just know it (ELL – Advanced)*
Comment 13 (0f 16): *The first one because it was so simple (non-ELL)*

Both of the students that circled a word or symbol still answered the question correctly even if they did not completely understand the mathematic technical terms *congruent* and *corresponding*, and the notation symbol for *segment* (as seen in $\overline{LM}$).

Table 11

Item Indicated as Easiest by non-ELLs and ELLS

	No Choice	Item 1 G/Visual	Item 2 D/No Visual	Item 3 D/Visual	Item 4 A/No Visual	Item 5 N/No Visual	Grand Total
Non-ELL	1	6	2	0	1	0	10
ELL	0	4	0	1	0	1	5
Grand Total	1(6%)	10 (63%)	2 (13%)	1(6%)	1 (6%)	1(%)	16

Note: G = Geometry Strand; Visual = Non-linguistic Feature; D = Data and Probability Strand; A = Algebra strand; N = Numbers and Operations Strand

Table 12

Item Indicated the Most Difficult by non-ELLs and ELLS

	No Choice	Item 4 A/No Visual	Item 5 N/No Visual	Item 4 and 5[1]	Grand Total
non-ELL	1	3	5	1	10
ELL	0	3	3	0	6
Grand Total	1 (6%)	6 (38%)	8 (50%)	1 (6%)	16

Note: Visual = Non-linguistic Feature; A = Algebra strand; N = Numbers and Operations Strand
[1]Participant chose both 4 and 5 as the most difficult

Item Two. This item had the highest number (6) of circled words and phrases, and the only one circled by an ELL. Four students answered the item incorrectly, 3 were ELLs and 1 was a non-ELL. Two students chose this answer as the easiest item (see Tables 10 and 11 for the easiest and most difficult items chosen by group); their comments explaining their choice follow:

Comment 12 (of 16): *2 because we did it already (non-ELL)*
Comment 14 (of 16): *Number 2 because you can do that way better than everything (non-ELL)*

No participant chose this item as being the most difficult. The ELL (low-intermediate) circled the entire phrase *six-sided number cube.* Each word in the phrase is context-specific vocabulary. The participant may be familiar with each word individually, but when strung together they are an enigmatic reference to a “die”.

This phrase is the commonly used term for a die in mathematics classrooms and textbooks, but could be very confusing to anyone that has not had experience with this phrase in their mathematics register. A cube has six faces or *sides* to begin with; this redundancy may also add to the confounding. This confusion could be easily be attenuated with the presence of a picture

The other words identified (all by non-ELLs) as unfamiliar or confusing were *numeral* (4 times) and *numerals* (1 time). *Number* is the more commonly used term, as it is the thing that *numerals* (a low frequency technical term) symbolize. Both *number* and *numeral* are used in the same item which may confuse the relationship between the two terms. *Number* could easily have been substituted for *numeral* without changing the construct, unless the item intended to measure the understanding of the word *numeral*.

Item Three. Over half of the participants answered this item incorrectly, yet none of the participants indicated that it was the most difficult problem for them. One participant, an ELL, denoted this item as the easiest, although they answered it incorrectly. This item included a chart with the necessary information to answer the problem. The term *relative frequency* was circled 4 times. It seems that many other participants did not understand this term either because they did not get the problem correct. Both *relative* and *frequency* are context-specific vocabulary words, but when paired together to form a compound noun they become a technical mathematics term specifically used in the Statistics branch of mathematics. A student that has little experience with the mathematical term may confuse the polysemous noun *relative* with its more commonly used meaning. Unless the construct is measuring a student's understanding of the term *relative frequency,* the more commonly used term *percent* could be substituted.

The other word circled in the item was the commonly used past tense verb *chose* from the question: What is the relative frequency of students who *chose* red as their favorite color*?* The participant was a non-ELL that only spoke English at home. The present tense of the same verb, *choose,* is also used in the item which may have been confusing to the participant.

Item Four. This was the only item on which ELLs outperformed non-ELLs. One participant indicated that this item was the easiest, while 7 (44%) of the participants thought it was the most difficult. A sample of their comments follow:

Comment 7 (of 16) – Easiest: *Number 4 because there were hints in the story (non-ELL, answered incorrectly)*
Comment 4 (of 16) – Difficult: *I would think #4 because its harder and I really didn't get it at all because the words I don't get what they mean (ELL, answered correctly)*
Comment 11 (of 16) – Difficult: *Number 4 was the hardest because I didn't understand all the math terms(ELL, answered correctly)*
Comment 12 (of 16) – Difficult: *4, I don't get temperature (non-ELL, answered incorrectly)*
Comment 16 (of 16) – Difficult: *#4 because it was confusing(non-ELL, answered incorrectly)*

The participants that circled any words or numbers and symbols were all non-ELLs that speak only English at home. It is clear from the comments above that other participants did not understand all of the words in the problem but they did not circle any of them. Two participants circled the word *Celsius,* one of which also circled -*13ºC* and *6 ºC. Celsius* is the scale and unit measurement of temperature commonly used throughout the world except for the United States, which uses Fahrenheit. This cross-discipline academic vocabulary word is used starting in 3rd grade mathematics and science classes, according to the county curriculum. Considering these students have attended school in the U.S. since kindergarten, they must have been exposed to the term previously but may have trouble seeing it in print and are more familiar with

Fahrenheit because it is more commonly used in the U.S. A future study could be conducted on this anomaly by changing the word *Celsius* to *Fahrenheit* to see if the performance of non-ELLs were negatively affected by the more culture specific term *Celsius.* A more plausible reason that more students answered incorrectly could be the increased rigor of the item, requiring students to translate the word problem into an algebraic equation.

A third participant circled the phrase *represents the Celsius* from the question: Which of the following best *represents the Celsius* temperature *n* hours after 1:00 p.m. on Tuesday? The participant may have understood most words in the phrase separately, as the transitive verb *represents* and the article *the* are both high frequency words and *Celsius* is used across disciplines. The whole phrase together asks the student for an abstract concept, which may have resulted in the incorrect answer. The participant had an E for their math grade from the previous trimester, so they may struggle with the abstract concepts present in algebra.

Item Five. Ten (62%) out of the 16 participants answered this item incorrectly, yet only one participant circled anything in the item. Over half of the students indicated that this item was the most difficult. A sample of their comments follow:

Comment 1 (of 16): *#5 because I didn't know how to do fractions very well. (ELL – answered incorrectly)*
Comment 2 (of 16): *Number 5 was the most difficult because the numbers they gave didn't tell you what to do with them. (non-ELL – answered correctly)*
Comment 8 (of 16): *The most difficult one was the mpg one because you have to think it through and decide where to start or where to finish.(non-ELL- answered correctly)*
Comment 11 (of 16): *Number 5 The desmines messed me up. (ELL- answered incorrectly)*

A common theme in the comments above is that the use of fractions was the main cause of incorrect answers. Comment 11 uses the term *desmines* which I interpreted to mean “decimals,” another term besides fraction used for a rational number. The participant that made Comment 2 recognized that this item assumed that students would know the relationship of miles per gallon without explicitly stating it. This concept could potentially be difficult because none of these students drive, and their family may not even own a car. That said, the trouble seems to be the use of fractions, which were the only features (along with the number 25) circled in the item. The student may have also circled those features out of habit because they are the numbers needed to solve the problem, a strategy that these students have been using.

Strategies participants use on test items with confusing and unfamiliar words. The last short answer question on the questionnaire regarded strategies that the participant uses when they don’t understand some of the words in a math item (see Appendix for all responses). Five (1 ELL, 4 non-ELLs) out of the 16 participants said they guess. Three others (1 ELL, 2 non-ELLs) try to use context clues. Asking the teacher was the strategy mentioned by three (1 ELL, 2 non-ELLs) more participants, although this is not allowed on high stakes tests. Other responses included: blanking out (ELL), skipping the problem and coming back to it later (non-ELL), and trying to remember if the word was ever mentioned in class (ELL). None of the participants mentioned using a glossary which the reviewed research identified as the most effective accommodation. This could be due to the fact that students rarely use a glossary in the classroom, let alone on high stakes tests.

Summary

This study used item performance data from the 7th and 8th grade 2007 and 2008 MEAP assessments and questionnaires to answer the research questions.

Logistic regression procedures conducted on the item performance data showed that the first two research questions had very significant (p<.0001) findings in that the ratio in the odds of passing an item number for ELLs and non-ELLs is affected by both whether that item is a computation or word problem, and also if a non-linguistic feature is present. A second analysis conducted with logistic regression procedures found that the ratio in the odds of passing an item for ELLs and non-ELLs is not affected by the GLCE Strand of the item, even whether it is Data and Probability or not. These findings answer the third research question.

Descriptive, cross tabulation, frequency, and content analyses were conducted on questionnaire responses from 16 participants. Descriptive analyses showed that overall non-ELLs outperformed ELLs on the word items, with ELLs outperforming non-ELLs on the item based on the Algebra Strand with no linguistic feature present. An analysis of the confusing or unfamiliar terms circled by participants showed that the majority were context-specific or technical mathematics language features which are all construct relevant. Some of the more technical words could be replaced with more commonly known words of phrases with the same meaning. Only one ELL circled anything, although others indicated that they did not understand all of the words in their comments. Further research, ideally think-alouds, need to be conducted in this area to gain a better understanding. Lastly, it was found that participants utilize a variety of strategies when they encounter unknown or unfamiliar words on math test items none of which were using a glossary, the most effective accommodation found in the research.

Chapter Five: Conclusion

Summary of the Study

English language learners (ELLs) must take the mathematics portion of state standardized tests regardless of their time spent in U.S. schools. Yet, they do not have to take the English language arts portion until they have in U.S. schools for at least a year. This practice follows the misconception that mathematics is a 'universal language' and less language dependant, however, a significant performance gap between non-ELLs and ELLs on high stakes mathematics tests persists and must be addressed.

Mathematics has its own complex register including academic vocabulary, symbols, numbers, and non-linguistic features (Brown, 2005; Lager, 2006; Wolf & Leon, 2009; Wright & Li, 2008), which learners must negotiate between in order to make meaning. Research has shown that it may take ELLs 5 – 7 years to acquire cognitive academic language proficiency (Cummins, 1981), and some test items are written at an above-grade reading level (Carter & Dean, 2006); therefore, the high stakes mathematics tests are a measure of both language proficiency, and mathematics knowledge and ability. As a result, the scores cannot be interpreted validly and because the consequences are so dire for schools with below-proficient scores more research is necessary to address this issue.

This study aimed to answer the central research question: To what extent does language proficiency influence achievement on high stakes mathematics tests? This broad question was broken down further into four sub-questions:

1. Do English language learners perform differently on computation questions as opposed to word problems on high stakes mathematics tests? How does their achievement compare to non-English language learners on these tests?

2. Is the difficulty of an item, due to language complexity, alleviated with the inclusion of graphical, pictorial, or schematic representations?

3. Are there specific strands, as defined by the Michigan Curriculum Framework (Data & Probability, Algebra, Number & Operations, Geometry, Measurement), that English language learners perform differently on in comparison to non-English language learners on high stakes test items?

4. What language issues do students identify as confusing or unfamiliar in high stakes math test items?

In order to answer these questions, a cross-sectional quantitative study was conducted. The study included item performance data, by group, for 24,693 seventh and eighth grade students who took the 2007 and/or the 2008 mathematics MEAP assessment, and a questionnaire completed by 16 seventh and eighth grade participants for triangulation. The item performance data set was analyzed using a logistic regression model to determine the interaction effects between ELLs and non-ELLs based on item type, item language, and item strand. Cross tabulation, content, descriptive ,and frequency analyses were conducted on the questionnaire responses and later compared to the results of the logistic regression analysis.

Findings from this study show that the ratio in the odds of passing an item for ELLs and non-ELLs is affected by both whether that item is a computation or word problem, and also if a non-linguistic feature is present with $p <.0001$. The odds ratio for non-ELLs compared to ELLs passing an item is smallest when the item is a computation problem including a non-linguistic feature. The odds ratio is greatest when the item is a word problem without a non-linguistic feature. Results from a second logistic regression analysis show that the difference in passing rate for non-ELLs and ELLs is not affected by the GLCE Strand the item is based on.

Non-English language learners outperformed ELLs on the five MEAP items included in the questionnaire. The majority of the 19 words identified as confusing or unfamiliar were context-specific or technical mathematics language features, only one of which was circled by an ELL. The participants indicate that they utilize a variety of strategies when they encounter unknown words but none mention the use of a glossary, which is the most effective accommodation according to the research.

Conclusions

The purpose of this study was to investigate the extent to which language proficiency influences achievement on mathematics high stakes tests. This study goes beyond current research by investigating the performance of ELLs on items categorized as computation or word problems rather than by level of linguistic complexity, and by using data from Michigan, a state with a growing population of ELLs. Furthermore, the study adds to prior research by Martiniello (2009) as suggested, which investigated of the influence of learning strands and the multi-semiotic representation of the language of mathematics on high stakes test performance . The hypothesis that language proficiency does influence performance is supported by the results of the study, and the extent of which is described in the following answers to the four research sub-questions.

In regard to the first questions, the difference in the passing rate on MEAP items is very significant between ELLs and non-ELLs, with non-ELLs outperforming ELLs. Furthermore, the likelihood of an ELL passing a computation problem is greater than that of passing a word problem; the same is true for non-ELLs.

The second question is answered by analyzing the interaction effects between item type and inclusion (or exclusion) of a non-linguistic feature. The impact of language complexity is alleviated for ELLs when a non-linguistic feature is included in the test item. The odds that a non-ELL will pass a computation or word problem over an ELL are reduced when a non-linguistic feature is included, meaning that the performance gap shrinks.

Additional analyses were conducted to answer the third question, which reveal that the difference in the passing rate of an item by a ELLs and non-ELLs is not based on the GLCE Strand of the item. Likewise, the difference in passing rates on items specifically based on the Data and Probability Strand were found to be insignificant. So, no particular GLCE Strand was shown to consistently and negatively affect ELL performance per item.

Lastly, the fourth question was answered through an analysis of questionnaire responses. The majority of the language features identified by participants as problematic were academic vocabulary, more specifically, context-specific and technical mathematics terms. However, most of these were identified by non-ELLs as only one ELL identified a language feature as being problematic.

Discussion

Results from this study reflect Cummin's (1981) theories of BICS and CALP, as ELLs are expected to acquire general English and mathematics academic proficiency and be able to translate between their own native language, English, and mathematics in order to make meaning of mathematics texts and problems. This issue is exacerbated on high stakes tests because the cognitively demanding items are context reduced and there is no feedback. In this section each result will be discussed in terms of its agreement or disagreement with the results from prior research studies, followed by a delineation of the educational implications.

Overall, this study provides support to the available research indicating that language proficiency does have an effect on the performance of ELLs on high stakes mathematics tests (Abedi et al., 2005; Abedi & Lord, 2001; Lager, 2006; Martiniello, 2008, 2009; Shaftel, Belton-Kocher, Glasnapp, & Poggio, 2006; Wolf & Leon, 2009). Previous research investigated the influence of language proficiency on performance using either a wide range of test items all loosely described as word problems and rated by level of linguistic complexity or word count, or solely traditional word problems. This study made a distinction between the types of problem to investigate further differences. non-ELLs were shown to outperform ELLs on both word and computation problems, although the gap was smaller on computation problems. This means that ELLs performed better on computation problems, which by definition are generally shorter in item length than word problems, which is consistent with prior research (Shaftel et al., 2006).

The inclusion of a non-linguistic feature alleviated the impact of linguistic complexity on both word and computation problems for ELLs. This was attested in the results showing that the odds a non-ELL will pass an item over an ELL is lessened by the presence of a non-linguistic feature. Although this study combined all types of non-linguistic features into one category, the results are similar to those from studies by Martiniello (2009) and Wolf and Leon (2009). Despite the fact that the inclusion of a non-linguistic feature could be an effective accommodation for ELLs, it may not be valid because the performance of non-ELLs also improved.

In contrast to both studies by Martiniello (2008, 2009), items based on the Data and Probability Strand did not significantly give favor to non-ELLs. Likewise, no GLCE Strand was found to significantly give favor to non-ELLs or ELLs. Teachers may have spent more time on the Data and Probability Strand, or it could be due to the fact that only 11 out of the 104 questions from the MEAP test data analysis were based on the Data and Probability Strand. Results from the questionnaire were also mixed, although the sample size and question bank were small.

As can be seen from the questionnaire results and prior research (Shaftel et al., 2006; Kieffer et al., 2009), the language features found by both ELLs and non-ELLs to be confounding were construct relevant academic vocabulary words and phrases. Although these language features are to be expected on a measure of mathematical knowledge and ability, some of the more technical words or phrases could be replaced with more common synonymous words or phrases to reduce confusion; that is, unless the questions was designed to measure knowledge of that word or phrase.

An interesting observation of note is that the term *Celsius* may have caused some non-ELLs to answer a question incorrectly. This term for temperature measurement is used by the majority of the world except for the U.S., and even though it is commonly used in science and math classes in U.S. classrooms, it still could have been a result of reverse-cultural bias. This type of bias is similar to examples mentioned by Brown (2009), adding to the evidence that mathematics is not as universal as traditionally thought.

The sense of ‘false knowing’ observed by Lager (2006) in his study of mathematics-language reading interactions was corroborated with the results of this study. Participants indicated struggling with some of the terminology, yet did not circle any words or phrases. Participants incorrectly answered problems that they deemed to be the easiest perhaps because they did not fully understand all of the terms and non-linguistic features. Other participants indicated struggling with some of the terminology, yet did not circle any words or phrases.

Implications. Results from this study have important implications for classroom instruction, test design and score interpretation. Since ELLs must acquire both the academic English register and the mathematics register simultaneously, teachers must be prepared to act as both a language and content area teacher.

Test designers and developers must make every effort to reduce linguistic complexity and cultural bias. Additionally, there is a need to develop a framework for valid accommodations that are appropriate for ELLs, instead of using the same framework designed specifically for special education students.

Lastly, test results must be approached with caution as the scores for every subgroup cannot be validly interpreted. Schools are being judged wrongfully for scores that are invalid measures of the intended constructs.

Recommendations

As the number of ELLs included in mainstream classrooms increase, the responsibilities of educators, administrators, and everyone involved in test design and decision making, must be expanded to ensure fair and equitable educational practices for students learning the English language. Currently, many teachers do not have much, if any, training for effectively serving ELLs in the mainstream classroom. Teacher preparation programs in colleges and universities should require a course in second language acquisition theory and methods for teaching ELLs. Furthermore, administrators should offer and require staff members to engage in quality professional development on instructing ELLs.

Specifically in the mathematics classroom, teachers should involve students in mathematical discourse with speaking, listening, reading, and writing through scaffolded instruction. Students should be engaged in cognitively demanding problem solving tasks that involve communication using the mathematics register with guidance and support. This practice will not only benefit ELLs , but all students who also need more experience engaging with the mathematics register and academic vocabulary development.

Legislation that attaches devastating consequences to schools based solely on high stakes test scores should cease until test design and score interpretation can be reexamined in order to give a fair picture of all students' abilities and knowledge. First of all, the MEAP assessment should also be analyzed for DIF between ELLs and non-ELLs, instead of just between male and females, and black and white students. Technical mathematics terms that are not being assessed explicitly should be replaced with more common synonymous content-specific terms. Although more research is needed, the inclusion of non-linguistic features on mathematics high stakes test items may reduce the impact of linguistic complexity.

This study utilized a cross-sectional design, only allowing for a snapshot of a point in time. Future studies should include longitudinal designs in order to gain a better picture of the issue over a longer period of time. This type of study could track individual student performance as time passes and English proficiency increases. Classroom performance could be included with standardized test scores for a more accurate picture.

Another limitation to this study is number of ELLs (7%) compared to the number of non-ELLs (93%) included in the MEAP item performance data set. This is especially true when looking at schools individually, because the values might reflect just one or two students. However, since 17 schools were included providing many data points and the demographics are very similar to the population, the bias is reduced. Future research should also include within-group investigation to account for the level of English proficiency and students who are former ELLs. This could provide more specific insight into how much student performance is influenced by different language proficiency levels.

Moreover, the questionnaire only included five assessment items and was administered to a small group of 16 participants. A larger group of participants would have been ideal, as well as, a larger bank of assessment items. Two versions of the questionnaire could be given: one with all word problems without non-linguistic features and one with the exact same word problems but each includes a non-linguistic feature. Given more time, future studies could incorporate these suggestions.

Finally, this study used convenience samples for data collection. Future studies should employ randomized sampling procedures so that the results can be generalized to a larger population. More grade levels could be included to expand the range of the results and implications. Any additional research will lead to more informed decision making in the realm of high stakes testing which is necessary for all students to be treated equally and fairly.

References

Abedi, J. (2002). Assessment and accommodations of English language learners: Issues, concerns, and recommendations. *Journal of School Improvement, 3*(1), 83-89.

Abedi, J. (2004). The no child left behind act and English language learners: Assessment and accountability issues. *Educational Researcher, 33*(1), 4-14.

Abedi, J., Bailey, A., Butler, F., Castellon-Wellington, M., Leon, S., & Mirocha, J. (2005). *The validity of administering large-scale content assessments to English language learners: An investigation from three perspectives.* (CSE report 663). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Abedi, J., Courtney, M., & Leon, S. (2003). *Research-supported accommodation for English language learners in NAEP.* (CSE technical report No. CSE-TR-586) Los Angeles: UCLA, Center for the Study of Evaluation.

Abedi, J., Courtney, M., Leon, S., Kao, J., & Azzam, T. (2006). *English language learners and math achievement: A study of opportunity to learn and language accommodation.* (Technical Report No. 702). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research, 74*(1), 1-28.

Abedi, J., Hofstetter, C., Baker, E., & Lord, C. (2001). *NAEP math performance and test accommodations: Interactions with student language background.* (CSE technical report No. CSE-TR-536). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing, Center for the Study of Evaluation, Graduate School of Education and Information Studies.

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

Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on english language learners' test performance. *Educational Measurement: Issues and Practice, 19*(3), 16-26.

Abedi, J., Lord, C., and Plummer, J. (1997). Final report of language background as a variable in NAEP mathematics performance (CSE Technical Report 429). Los Angeles: University of California, Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.

Barbu, O. C., & Beal, C. R. (2010). Effects of linguistic complexity and math difficulty on word problem solving by English learners. *International Journal of Education*, *2*(2). Retrieved January 20, 2011, from http://www.macrothink.org/journal/index.php/ije/article/viewFile/508/372

Beal, C. R., Adams, N. M., & Cohen, P. R. (2010). Reading proficiency and mathematics problem solving by high school English language learners. *Urban Education, 45*(1), 58-74.

Brown, C. (2005). Equity of literacy-based math performance assessments for english language learners. *Bilingual Research Journal, 29*(2), 337-363.

Brown, C. L., Cady, J. A., & Taylor, P. M. (2009). Problem solving and the English language learner. *Mathematics Teaching in the Middle School, 14*(9), 532-539.

Carter, T. A., & Dean, E. O. (2006). Mathematics intervention for grades 5-11: Teaching mathematics, reading, or both? *Reading Psychology, 27*(2-3), 127-146.

Crawford, J. (1990). Language freedom and restriction: A historical approach to the official language controversy. In J. Reyhner (Ed.), *Effective Language Education Practices & Native Language Survival* (pp. 9-22). Choctaw: Native American Language Issues. Retrieved January 25, 2011, from http://jan.ucc.nau.edu/~jar/NALI2.html

Crawford, J. (2004). No child left behind: A misguided approach to school accountability for English language learners. In *National Association for Bilingual Education*. Retrieved January 11, 2011, from www.nabe.org/files/NABE_on_NCLB.pdf

Crawford, J. (2009). No child left behind: A failure for English language learners. In *Institute for Language and Education Policy*. Retrieved January 21, 2011, from http://www.elladvocates.org/documents/nclb/NCLB_and_ELLs.pdf

Cuevas, G. J. (1984). Mathematics learning in English as a second language. *Journal for Research in Mathematics Education, 15*(2), 134-144.

Cummins, J. (1981) Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics,* 2, 132-149.

Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire.* Clevedon, England: Multilingual Matters Ltd.

Echevarria, J., Short, D., & Powers, K. (2006). School reform and standards-based education: A model for English-language learners. *Journal of Educational Research, 99*(4), 195-210.

Gay, L. R., Mills, G. & Airasian, P. W. (2009). Educational Research: Competencies for Analysis and Interpretation. (9th edition). NJ: Pearson.

Hansen-Thomas, H. (2009). Reform-oriented mathematics in three 6th grade classes: How teachers draw in ELLs to academic discourse. *Journal of Language, Identity & Education, 8*(2), 88-106.

Haycock, K. (2006). No more invisible kids. *Educational Leadership, 64*(3), 38-42.

Hofstetter, C. H. (2003). Contextual and mathematics accommodation test effects for English-language learners. *Applied Measurement in Education, 16*(2), 159-188.

Janzen, J. (2008). Teaching English language learners in the content areas. *Review of Educational Research, 78*(4), 1010-1038.

Johnson, E., & Monroe, B. (2004). Simplified language as an accommodation on math tests. *Assessment for Effective Intervention, 29*(3), 35-45.

Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research, 79*(3), 1168-1201.

Kindler, A.L., "*Survey of the states' limited English proficient students and available educational programs and services, 2000-2001 Summary Report,"* National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs.

Lager, C. A. (2006). Types of mathematics-language reading interactions that unnecessarily hinder algebra learning and assessment. *Reading Psychology, 27*(2-3), 165-204.

Lamb, J. H. (2010). Reading grade levels and mathematics assessment: An analysis of Texas mathematics assessment items and their reading difficulty. *Mathematics Educator, 20*(1), 22-34.

Lemke, J. (2003). *Mathematics in the Middle: Measure, picture, gesture, sign, and word*. Retrieved February 3, 2011, from http://academic.brooklyn.cuny.edu/education/jlemke/papers/myrdene.htm

MacSwan, J. (2000). The threshold hypothesis, semilingualism, and other contributions to a deficit view of linguistic minorities. *Hispanic Journal of Behavioral Sciences, 22*(1), 3-45.

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

Martiniello, M. (2009). Linguistic complexity, schematic representations, and differential item functioning for English language learners in math tests. *Educational Assessment, 14*(3-4), 160-179.

Menken, K. (2008). *English Learners Left Behind: Standardized Testing as Language Policy* (pp. 1-31). Clevedon: Multilingual Matters.

Michigan Department of Education: Office of Assessment and Accountability, Pearson Educational Measurement, and Harcourt Assessments, Inc.(n.d.). Michigan Educational Assessment Program: Technical Report 2008-2009. In *OEAA Technical Reports*. Retrieved February 22, 2011, from http://www.michigan.gov/mde/0,1607,7-140-22709_35150-172665--,00.html

Michigan Department of Education: Office of School Improvement. (2006). Mathematics Grade Level Content Expectations v. 12.05. In *K-8 Complete GLCE Mathematics Documents*. Retrieved February 6, 2011, from http://michigan.gov/documents/MathGLCE_140486_7.pdf

Michigan Department of Education (2007). MEAP Released Items: Grade 7, 8. In *Michigan Educational Assessment Program.* Retrieved February 6, 2011, from http://www.michigan.gov/documents/mde/F07_Gr8_Math_Rel_Doc_222005_7.pdf http://www.michigan.gov/documents/mde/F07_Gr7_Math_Rel_Doc_222003_7.pdf

Michigan Department of Education (2008). MEAP Released Items: Grade 7, 8. In *Michigan Educational Assessment Program.* Retrieved February 6, 2011, from http://www.michigan.gov/documents/mde/F08_Gr8_Math_Rel_Doc_269514_7.pdf http://www.michigan.gov/documents/mde/F08_Gr7_Math_Rel_Doc_269513_7.pdf

Michigan Department of Education. (2010). Test Administrator Manual: Grade 8. In *Michigan Educational Assessment Program*. Retrieved February 6, 2011, fromhttp://www.michigan.gov/documents/mde/MEAP10_DFA_G8_FNL_329978_7.pdf

Michigan Department of Education. (2010). Test Administrator Manual: Grade 8. In *Michigan Educational Assessment Program*. Retrieved February 6, 2011, from http://www.michigan.gov/documents/mde/MEAP10_DFA_G8_FNL_329978_7.pdf

Moschkovich, J. (2007). Using two languages when learning mathematics. *Educational Studies in Mathematics, 64*(2), 121-144.

Ockey, G. J. (2007). Investigating the validity of math word problems for English language learners with DIF. *Language Assessment Quarterly, 4*(2), 149-164.

Office of Educational Assessment and Accountability. (2009). *Assessment Accommodation Summary Table*. Retrieved February 9, 2011, from http://www.michigan.gov/documents/mde/Updated_Revised_Accommodation_Summary_Table_101110_334848_7.pdf

O'Halloran, K. L. (2000). Classroom discourse in mathematics: A multi-semiotic analysis. Linguistics and Education, 10(3), 359–388.

Rivera, C., & Collum, E. (2004). *An analysis of state assessment policies addressing the accommodation of English language learner.* Retrieved from the National Assessment Governing Board website: http://www.nagb.org

Robinson, J. P. (2010). The effects of test translation on young English learners' mathematics performance. *Educational Researcher, 39*(8), 582-590.

Sato, E., Rabinowitz, S., Gallagher, C., & Huang, C. (2010). *Accommodations for English language learner students: The effect of linguistic modification of math test item sets.* (Report No. NCEE 2009-4079) Retrieved from the National Center for Education Evaluation and Regional Assistance website: http://ies.ed.gov/ncee/.

Shaftel, J., Belton-Kocher, E., Glasnapp, 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.

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher, 32*(2), 3-13.

Teachers of English to Speakers of Other Languages, Inc. (1997). *ESL standards for pre-*K–12 students. Alexandria, VA: Author.

U.S. Department of Education, Office of Elementary and Secondary Education. (2007). Assessment and Accountability of Recently Arrived and Former Limited English Proficient (LEP) Student: Non-Regulatory Guidance. Retrieved from www2.ed.gov/policy/elsec/guid/lepguidance.doc

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

Wall, D., & Alderson, J. C. (1993). Examining washback: The sri lankan impact study. *Language Testing, 10*(1), 41-69.

Willner, L. S., Rivera, C., & Acosta, B. D. (2009). Ensuring accommodations used in content assessments are responsive to English-language learners. *Reading Teacher, 62*(8), 696-698.

Wolf, M. K., Kim, J., Kao, J. C., & Rivera, N. M. (2009). Examining the effectiveness and validity of glossary and read-aloud accommodations for English language learners in a math assessment. (CRESST report 766).

Retrieved from the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) web site: http://www.cresst.org.

Wolf, M. K., & Leon, S. (2009). An investigation of the language demands in content assessments for English language learners. Educational Assessment, 14(3-4), 139-159.

Wright, W. E., & Li, X. (2008). High-stakes math tests: How no child left behind leaves newcomer English language learners behind. *Language Policy, 7*(3), 237-266.

Young, J. W. (2009). A framework for test validity research on content assessments taken by English language learners. *Educational Assessment, 14*(3-4), 122-138.

Appendix A

Permission to Conduct Research at the Middle School

Principal Consent Form

I. Research Background (to be completed by researcher)

Title of the Study To What Extent Does Language Proficiency Influence Achievement on High Stakes Tests?

Name of Researcher Debbie Schuitema Phone [redacted]

Street add [redacted]

E mail: [redacted]

II. Description of Research Proposal

Researcher is to provide the principal with a copy of the executive summary and the time requirement form.

III. Agreement (to be completed by principal)

I, ________, principal of ________ school, understand

- that study and what it requires of the staff, students, and/or parents in my school,
- that the privacy and confidentiality of any staff or student will be protected,
- that I have the right to allow or reject this research study to take place at my school,
- that I have the right to terminate the research study at any time,
- that I have the right to review all consent forms and research documents at any time during the study and up to three years after the completion of the study.

☒ I grant permission to the researcher to conduct the above named research in my school as described in the proposal.

I DO NOT grant permission to the researcher to conduct the above named research in my school as described in the proposal.

☐ I understand that data should be released only by the departments that own them. My staff and I shall not release data to the researcher without approval from the IRB.

____

Signature of Principal

Appendix B

Questionnaire in English and Spanish

Thank you for taking the time to participate in this study.

This questionnaire is part of a research project that is studying how different features of test questions may affect student performance. Your answers will help anyone involved in test design to create fair questions for all students.

This questionnaire is divided into three sections: math problems, questions about the math problems, and demographic information.

You may use a calculator on all of the math problems. Please complete each section and do NOT write your name anywhere on this paper.

You may raise your hand at any time if you have any questions, and a teacher will come to assist you.

You will NOT be graded on your answers and this test will NOT be a part of your grade for this class.

It should take about 15 minutes to complete this questionnaire.

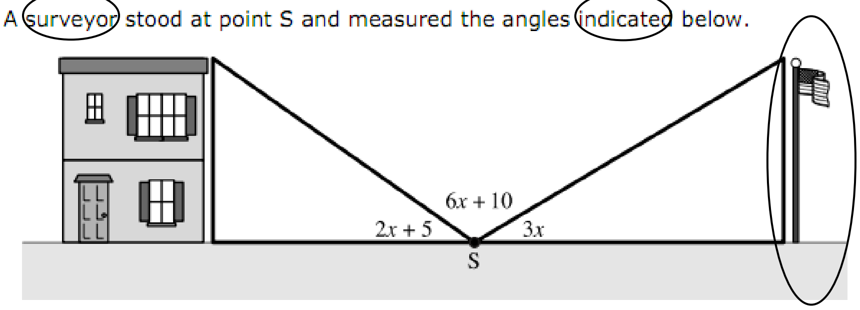
When you are finished, please turn your questionnaire over and put your pencil down. Your teacher will collect it from you.

This research protocol has been approved by the Human Research Review Committee at Grand Valley State University. File No. 11-110-H Expiration: March 24, 2012.

Section One: Math Problems.
Please circle any confusing or unfamiliar words, phrases, or pictures as you answer the problems. Show your work (calculation, drawings) on each problem.

Example:

A surveyor stood at point S and measured the angles indicated below.



Which of the following equations could be used to find the number of degrees in each of these three angles?

Number of degrees in a straight line = 180 degrees.

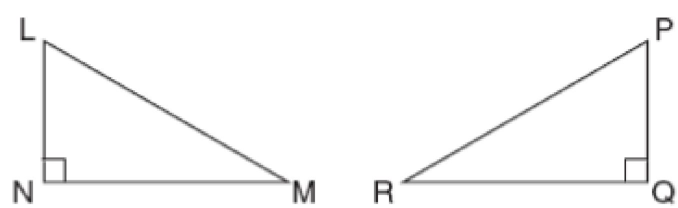
A $11x + 15 = 180$

B $11x + 50 = 180$

C $12x + 15 = 180$

D $36x + 50 = 180$

1. Triangle LNM is congruent to triangle PQR, as shown below.



What side of triangle PQR corresponds to $\overline{LN}$ in triangle LNM?

A $\overline{PQ}$

B $\overline{QR}$

C $\overline{RP}$

D $\overline{NM}$

2. Elizabeth is going to roll a fair six-sided number cube on which each face is labeled with a different numeral. If the numerals are 1 through 6, what is the probability she will roll a 3 on the first roll?

A $\frac{1}{2}$

B $\frac{1}{3}$

C $\frac{1}{5}$

D $\frac{1}{6}$

3. Mrs. Lee asked the 25 students in her class to choose their favorite color. The responses she received are in the table below.

Students' Favorite Color

Favorite Color	Number of Students
Red	5
Orange	2
Blue	7
Green	10
Purple	1

What is the relative frequency of students who chose red as their favorite color?

A 0.25

B 0.20

C 0.10

D 0.05

4. The temperature at 1:00 p.m. on Tuesday was $^{-}13$°C. There was an increase of 6°C per hour starting at 1:00 p.m. Which of the following best represents the Celsius temperature *n* hours after 1:00 p.m. on Tuesday?

A $^{-}13 + (6n)$

B $^{-}13 - (6n)$

C $(^{-}13n) + 6$

D $(^{-}13n) - 6$

5. A certain car can travel 25 miles on $2\frac{1}{4}$ gallons of gasoline. At this rate, which of the following is *closest* to the total number of miles the car can travel on $12\frac{1}{2}$ gallons of gasoline?

A 50

B 150

C 250

D 300

Section Two: Questions about the math problems.

1. Which problem was the easiest for you? What made it easier than the others?

2. Which problem was the most difficult for you? What made it more difficult than the others?

3. What do you do if you do not understand some of the words in a math problem on an important test?

Section Three: Demographic Information. Please complete all of the questions below as best as you can. If you do not know the answer to a question, please write "Not Sure".

1. *Please circle one:*

   I am a 7th grade female **I am a 7th grade male**

   I am an 8th grade female **I am an 8th grade male**

Please put the appropriate information on the line provided.

2. **Age**: ________

3. **Math grade from last trimester**: __________

4. *Please circle one.* **Are you an English language learner (ELL)?** :

   Yes No Not Sure

   If you answered yes, please circle the level:

   Beginner Low Intermediate High Intermediate Advanced Not Sure

For questions 5-6, please circle all that apply.

5. **We speak these languages in my home**:

   English Spanish Arabic Other (please list the language/s):

6. **I attended school in the U.S. for these grades:**

   Kindergarten 1st grade 2nd grade 3rd grade 4th grade

   5th grade 6th grade 7th grade 8th grade

Thank you so much for completing this questionnaire! Please turn your paper over so that your teacher knows you have finished.

¡Gracias por tomarse el tiempo para participar en este studio!

Este cuestionario forma parte de un proyecto de investigación que estudia cómo las características diferentes de las preguntas de la prueba pueden afectar el desempeño del estudiante. Sus respuestas ayudarán a cualquier persona involucrada en el diseño de prueba para crear las preguntas justas para todos los estudiantes.

Este cuestionario está dividido en tres secciones: los problemas de matemáticas, las preguntas sobre los problemas de matemáticas, y información demográfica.

Usted puede usar una calculadora en todos los problemas de matemáticas. Por favor completa cada sección y no escriba su nombre en cualquier parte de este trabajo.

Se puede levantar la mano en cualquier momento si usted tiene alguna pregunta, y un maestro vendrá a ayudarle.

NO se le califica en sus respuestas y esta prueba no será parte de la calificación de esta clase.

Debe tener alrededor de 15 minutos para completar este cuestionario.

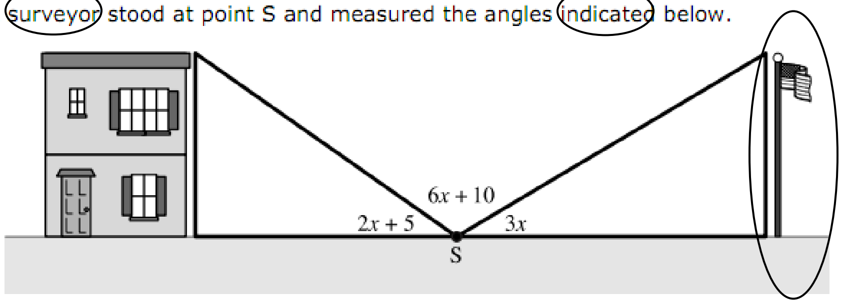
Cuando haya terminado, por favor, a su vez el cuestionario sobre la mesa. El maestro lo tomará.

Este protocolo de investigación ha sido aprobado por la Investigación en Seres Humanos del Comité de Examen de Grand Valley State University. Archivo Nº: 11-110-H

Section One: Math Problems.
Please circle any confusing or unfamiliar words, phrases, or pictures as you answer the problems. Show your work (calculation, drawings) on each problem.

Example:

A surveyor stood at point S and measured the angles indicated below.



Which of the following equations could be used to find the number of degrees in each of these three angles?

~~Number of degrees~~ in a straight line = 180 degrees.

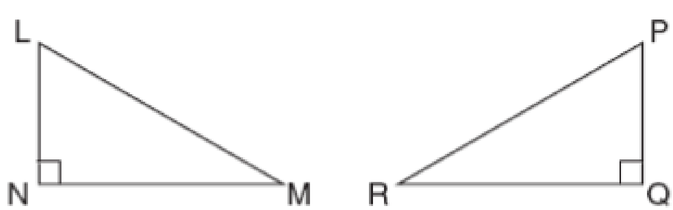
A $11x + 15 = 180$

B $11x + 50 = 180$

C $12x + 15 = 180$

D $36x + 50 = 180$

1. Triangle LNM is congruent to triangle PQR, as shown below.



What side of triangle PQR corresponds to $\overline{LN}$ in triangle LNM?

A $\overline{PQ}$

B $\overline{QR}$

C $\overline{RP}$

D $\overline{NM}$

2. Elizabeth is going to roll a fair six-sided number cube on which each face is labeled with a different numeral. If the numerals are 1 through 6, what is the probability she will roll a 3 on the first roll?

A $\frac{1}{2}$

B $\frac{1}{3}$

C $\frac{1}{5}$

D $\frac{1}{6}$

3. Mrs. Lee asked the 25 students in her class to choose their favorite color. The responses she received are in the table below.

Students' Favorite Color

Favorite Color	Number of Students
Red	5
Orange	2
Blue	7
Green	10
Purple	1

What is the relative frequency of students who chose red as their favorite color?

A 0.25

B 0.20

C 0.10

D 0.05

4. The temperature at 1:00 p.m. on Tuesday was $^{-}13$°C. There was an increase of 6°C per hour starting at 1:00 p.m. Which of the following best represents the Celsius temperature *n* hours after 1:00 p.m. on Tuesday?

A $^{-}13 + (6n)$

B $^{-}13 - (6n)$

C $(^{-}13n) + 6$

D $(^{-}13n) - 6$

5. A certain car can travel 25 miles on $2\frac{1}{4}$ gallons of gasoline. At this rate, which of the following is *closest* to the total number of miles the car can travel on $12\frac{1}{2}$ gallons of gasoline?

A 50

B 150

C 250

D 300

Sección Segunda: Las preguntas sobre los problemas de matemáticas.

1. ¿Cuál problema era más fácil para usted? Lo que hizo más fácil que los demás?

2. ¿Cuál problema era el más difícil para usted? Lo que hizo más difícil que los demás?

3. ¿Cuál hacer si usted no entiende algunas de las palabras en un problema de matemáticas en una prueba importante?

Sección Tres: Información Demográfica. Por favor, complete todas las preguntas a continuación lo mejor que puede. Si usted no sabe la respuesta a una pregunta, por favor, escriba "No estoy seguro".

1. *Por favor, marque uno:*:

Soy una mujer de 7 º grado　　　**Soy un hombre de 7 º grado**

Soy una mujer de 8 º grado　　　**Soy un hombre de 8 º grado**

Por favor, ponga la información apropiada en la línea provista.

2. **Edad**: ________

3. **Matemáticas grados del último trimestre**:__________

4. *Por favor, marque uno.* **Si eres un estudiante de idiomas Inglés(ELL**)? :

Sí　　　No　　　No estoy seguro

Si su respuesta es sí, por favor marque el nivel:

Principiante　　　Intermedio　　　Intermedio Avanzado
No Seguro

Para las preguntas 5-6, por favor marque todas las que correspondan.

5. **Se habla estos idiomas en mi casa**:

Inglés　　　Español　　　Árabe　　　Otros (por favor, indique el idioma / s):

6. **Asistí a la escuela en los Estados Unidos para estos grados**:

Jardín de niños　　1er grado　　2do grado　　3er grado　　4 º grado

5 º grado　　6 º grado　　7 º grado　　8 º grado

¡Muchas gracias por completar este cuestionario! Por favor, a su vez su papel más para que el maestro sabe que haya terminado.

Appendix C

HRRC Approved Protocol Letter

GRAND VALLEY STATE UNIVERSITY
www.gvsu.edu

DATE:	March 24, 2011
TO:	Deborah Schuitema
FROM:	Grand Valley State University Human Research Review Committee
STUDY TITLE:	[214560-3] Investigating the world of mathematics to uncover how language proficiency influences English language learners performance on high stakes tests
REFERENCE #:	11-110-H
SUBMISSION TYPE:	Revision
ACTION:	APPROVED
APPROVAL DATE:	March 24, 2011
EXPIRATION DATE:	March 24, 2012
REVIEW TYPE:	Expedited Review

Thank you for your submission of materials for this research study. The Human Research Review Committee has approved your revised research plan application as compliant with all applicable sections of the federal regulations, Michigan law, GVSU policies and HRRC procedures. All research must be conducted in accordance with this approved submission.

This approval is based on no greater than minimal risk to research participants. This study has received expedited review, category 2-7, based on the Office of Human Research Protections 1998 Guidance on Expedited Review Categories.

Please note that when the translated documents are available they should be uploaded to the protocol file on IRBNet as a new package. They should be accommpanied by an attestation statement from an appropriate authority re: veracity of translation.

Please insert the following sentence into your information/consent documents as appropriate. All project materials produced for participants or the public must contain this information.

> ***This research protocol has been approved by the Human Research Review Committee at Grand Valley State University. File No. 11-110-H Expiration: March 24, 2012.***

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note the following in order to comply with federal regulations and HRRC policy:

1. Any revision to previously approved materials must be approved by this office prior to initiation. Please use the *Change in Protocol* forms for this procedure. This includes, but is not limited to, changes in key personnel, study location, participant selection process, etc.

2. All UNEXPECTED PROBLEMS and SERIOUS ADVERSE EVENTS to participants or other parties affected by the research must be reported to this office within two days of the event occurrence. Please use the UP/SAE Report form.
   All instances of non-compliance or complaints regarding this study must be reported to this office in a timely manner. There are no specific forms for this report type.
3. All required research records must be securely retained in either paper or electronic format for a minimum of three years following the closure of the approved study. This includes signed consent documents from all participants.
4. This project requires continuing review by our office on an annual basis. Please use the appropriate *Continuing Review* forms when applying for approval extension.
   - Protocols that are active and open for enrollment require both the Primary Investigator and Authorizing Official to electronically sign the Continuing Review submission in IRBNet.
   - Protocols that are open for data analysis ONLY, require the Primary Investigator's signature.

If you have any questions, please contact the HRRC Office, Monday through Thursday, at (616) 331-3197 or hrrc@gvsu.edu. The office observes all university holidays, and does not process applications during exam week or between academic terms. Please include your study title and reference number in all correspondence with this office.

cc:

Appendix D

Parent Consent Forms in English and Spanish

Thesis Title: *Investigating the world of mathematics to uncover how language proficiency influences English language learners performance on high stakes tests*

Dear Parent or Guardian,

I am the math coach at Lee High School and a graduate student at Grand Valley State University in the Teaching English to Speakers of Other Languages program. I am currently doing my research on the extent to which language proficiency influences achievement on high stakes math tests. I am requesting your permission to learn more about your child's educational needs through a written questionnaire administered at school in English and/or the child's native language. The questionnaire will contain math problems similar to those found on the mathematics Michigan Educational Assessment Program (MEAP) test, and questions regarding the words and graphics used in those problems. Results from this study may inform and guide educators, administrators, legislators, test designers and parents in the assessment design, classroom instruction, and decision making that affect English language learners. As a result, your child or other students who are English learners may benefit from this study, if policy decision makers mentioned above take into consideration partially or entirely the recommendations of this study.

Your child's participation in this research study is completely voluntary. They do not have to participate and may quit at any time without any penalty. Please note that your child's name will not recorded on any of the information. The name of the school will be kept confidential and all data will only be used for the purpose of this study. At the end of the study, the data will be destroyed.

If you are willing for your child to participate, I would appreciate your signing and returning this letter to your child's math teacher in the next 2 days. If you have any questions about this study please contact me at Lee High School by phone at 616-452-4350 or email: dschuitema@godfrey-lee.org. You can contact the faculty advisor, Nagnon Diarrassouba, at 616-331-6611 or e-mail: diarrasn@gvsu.edu

If you have any questions about your child's rights as a research participant, please contact the Research Protections Office at Grand Valley State University, Grand Rapids, MI. Phone: 616-331-3197 or e-mail: HRRC@GVSU.EDU

Thank you for your cooperation.

Sincerely,

Debbie Schuitema
Math Coach – Center for Excellence in Education, CMU
Lee High/Middle School

This research protocol has been approved by the Human Research Review Committee at Grand Valley State University. File No. 11-110-H Expiration: March 24, 2012.

Check one:

☐ I want to participate and would like the questionnaire to be in (circle one):

English Spanish Arabic

☐ I do not want to participate

____

(student signature)

Parent/Guardian Permission

I give permission for my child, ___________________________, to participate in the questionnaire. (child's name)

____ ___________________________

(your signature) (Date)

GRAND VALLEY STATE UNIVERSITY
www.gvsu.edu

Título de la Tesis: *Investigación del mundo de las matemáticas para dar a conocer cómo el dominio del idioma influye el desempeño de los estudiantes del idioma inglés en los exámenes de alta importancia.*

Estimado Padre ó Guardián,

Soy la tutora de las matemáticas en la secundaria Lee High School y soy estudiante de post-grado en Grand Valley State University en el programa de Enseñanza del Inglés a Personas que Hablan otros Idiomas. Actualmente estoy haciendo mi investigación sobre la magnitud en la que el dominio del idioma influye el desempeño en exámenes de matemáticas de alta importancia. Estoy solicitando su permiso para conocer más acerca de las necesidades educativas de su niño, a través de un cuestionario escrito administrado en la escuela en inglés y/o en la lengua materna del niño. El cuestionario contendrá problemas de matemáticas similares a los que son encontrados en el examen de Matemáticas del Programa de Evaluación Educativa de Michigan (MEAP por sus siglas en inglés) y preguntas relacionadas con las palabras y gráficas usadas en esos problemas. Los resultados de este estudio pueden informar y guiar a los educadores, administradores, legisladores, diseñadores de exámenes y a padres en el diseño de evaluaciones, enseñanza de las clases y la toma de decisiones que afecta a los estudiantes del idioma inglés. Como consecuencia, su hijo u otros estudiantes que están aprendiendo inglés, pueden beneficiarse de este estudio, si los responsables de la decisión de las políticas mencionados anteriormente consideran parcial o totalmente las recomendaciones de este estudio.

La participación de su hijo en este estudio de investigación es totalmente voluntaria. No tiene que participar y puede abandonarlo en cualquier momento sin ningún castigo. Por favor tenga en cuenta que el nombre de su hijo no será documentado en ninguna información. El nombre de la escuela será mantenido confidencial y todos los datos serán utilizados solamente para el propósito de este estudio. Los datos serán destruidos al final del estudio.

Si usted está dispuesto a que su hijo participe, apreciaría su firma y la devolución de esta carta al maestro/a de matemáticas de su niño en los próximos 2 días. Si usted tiene alguna pregunta sobre este estudio, por favor contácteme por teléfono en la secundaria Lee High School al 616-452-4350 o por correo electrónico: dschuitema@godfrey-lee.org. Usted puede comunicarse con el consejero de la facultad, Nagnon Diarrassouba, al 616-331-6611 o por correo electrónico: diarrasn@gvsu.edu

Si usted tiene alguna pregunta sobre los derechos de su hijo como participante de la investigación, por favor comuníquese con la Oficina de Protecciones de Investigación en Grand Valley State University, Grand Rapids, MI. Teléfono: 616-331-3197 o correo electrónico: HRRC@GVSU.EDU

Gracias por su cooperación.

Atentamente,
Debbie Schuitema
Tutora de Matemáticas – Centro de Excelencia en Educación, CMU
Lee High/Middle School

Este protocolo de investigación ha sido aprobado por la Investigación en Seres Humanos del Comité de Examen de Grand Valley State University. Archivo Nº: 11-110-H Expiración: March 24, 2012.

Marque una:

☐ Yo quiero participar ☐ Yo no quiero participar

____
(Firma de estudiante)

Permiso de Padre/Guardián

Doy permiso para que mi niño/a, ___________________________, participe en el cuestionario.
(Nombre del niño/a)

____ ___________________________
(Su Firma) (Fecha)

Appendix E

Questionnaire Short Answer Responses

Questionnaire Short Answer Responses

1. Which problem was the easiest for you? What made it easier than the others?

1) Number One because it was easy to figure out the q and p thing
2) Number One was the easiest because I really understand triangles and the others were more difficult
3) I would think #5 I think it was easy to answer
4) 1 b/c I have learned it more often than others
5) 1 I knew how 2 do it
6) Number one was the easiest. Because you just had to flip it
7) Number 4 because there were hints in the story
8) The easiest problem was the triangle problem because you could easily see which part of the triangle matches
9) I'll say the graphing with the slope is easiest. By understanding it than the other
10) Number one was the easiest because I could visualize the problem
11) Number 1 I could see it and I just know it
12) 2 because we did it all ready
13) The first one because it was so simple
14) Number 2 because you can do that way better than everything
15) the one when the teacher asked who's favorite class was what
16) #1 it was more simple

2. Which problem was the most difficult for you? What made it more difficult than the others?

1) #5 because I didn't know how to do fractions very well
2) Number 5 was the most difficult because the numbers they gave didn't tell you what to do with them
3) I would think #4 because its harder and I really didn't get it at all because the words I don't get what they mean
4) 4 I am not used to new problems
5) 4 & 5 I didn't know how 2 do it
6) The fourth one was hard. Because you had to mutipliy
7) Number 5 because there were big fractions that were hard
8) The most difficult one was the mpg one because you have to think it through and decide where to start or where to finish
9) Some time I have difficulting with rise over run. Some time because I always the order mix up
10) Number 4 was the hardest because I didn't understand all the math terms

11) Number 5 The desmines messed me up
12) 4, I don't get temperature
13) The last one because it was confusing
14) It was number 5 because I didn't get it
15) The one with the car
16) #4 because it was confusing

3. **What do you do if you do not understand some of the words in a math problem on an important test?**
    1) I blank out and say I can't do it
    2) I always try to figure the words out and if I can't then I will try my hardest to get the right answer
    3) Like represents, reflected, increase or decrease
    4) I will try 2 remember if we have studied them or usually take my time 2 studie the answer
    5) I guess
    6) I ask the teacher
    7) You get up and ask the teacher for help
    8) I use other words in the sentence to see if I can connect to those words and see how they fit together
    9) No answer
    10) I just try to make my best guess
    11) I guess
    12) Try to make my best guess
    13) Skip the problem and come back 2 it later
    14) you ask a teacher for some help or try to remember how to do it
    15) try to get an easy word to replace it
    16) I guess and hope for the best

Appendix F

Letter to the Office of Educational Assessment and Accountability for Copyright Permission & E-mail Response

March 18, 2011

Michigan Department of Education
Office of Educational Assessment & Accountability
PO Box 30008
Lansing, MI 48909
oeaa@michigan.gov

To whom it may concern:
I am currently enrolled in the Grand Valley State University (GVSU) Graduate Studies in Education Program, and I am writing a thesis for the completion of my Master's in Education. My *thesis* is entitled "*Investigating the world of mathematics to uncover how language proficiency influences English language learners performance on high stakes tests.*" May I receive permission to include in the appendices, copies of the following items?

Five released items from the 2007 and 2008 7th and 8th grade mathematics MEAP Copyright © 2007, 2008, held by the Michigan Department of Education, State of Michigan

These items will also be included in a short questionnaire asking students to complete each problem as well as circle any unfamiliar or confusing words, or non-linguistic schematic representations. A sample questionnaire is attached.

Your signature at the bottom portion of this letter confirms your ownership of the requested items. The inclusion of your copyrighted material will not restrict your republication of the material in any other form. Please advise if you wish a specific copyright notice to be included on each page. My thesis will be cataloged in the GVSU library and will be available to other students and colleges for circulation.

Sincerely,
Deborah Schuitema
647 Lyon NE
Grand Rapids, MI 49503
deschuitema@gmail.com

PERMISSION IS GRANTED to you, Deborah Schuitema, to include the requested materials in her GVSU Master's of Education thesis.

Name of Company/Organization

Permission granted by: ________________________

Title: ________________________

Date: ________________________

Debbie Schuitema <deschuitema@gmail.com>

Graduate Student Permission Request

Viger, Steve (MDE) <VigerS@michigan.gov> **Thu, Apr 21, 2011 at 11:35 AM**
To: "deschuitema@gmail.com" <deschuitema@gmail.com>
Cc: OEAA <OEAA@michigan.gov>, "Keesler, Venessa (MDE)" <KeeslerV@michigan.gov>, "Griffiths, James (MDE)" <GriffithsJ@michigan.gov>

Hi Debbie:

Are you referring to items downloaded from the public OEAA site? If so they are fine to use as long as they are properly referenced as to where you found them.

--Steve

Steven G. Viger

Manager of Measurement Research & Psychometrics

Office of Accountability, Research, & Evaluation (OARE)

Office Phone: (517) 241-2334
Email: VigerS@michigan.gov

From: Griffiths, James (MDE)
Sent: Thursday, April 21, 2011 11:33 AM
To: deschuitema@gmail.com
Cc: OEAA; Viger, Steve (MDE); Keesler, Venessa (MDE)
Subject: RE: Graduate Student Permission Request

[Quoted text hidden]

GRAND VALLEY STATE UNIVERSITY
ED 693/695 Data Form

NAME: Deborah Schuitema

MAJOR: (Choose only1)

_____ Adult & Higher Ed	_____ Ed Differentiation	_____Library Media
_____ Advanced Content Spec	_____Ed Leadership	_____ Middle Level Ed
_____ Cognitive Impairment	_____ Ed Technology	_____Reading
_____ CSAL	_____ Elementary Ed	_____School Counseling
_____ Early Childhood	_____ Emotional Impairment	_____SecondaryLevel Ed
_____ ECDD	_____ Learning Disabilities	_____ Special Ed Admin
		___x_ TESOL

TITLE: Investigating the World of Mathematics to Uncover How Language Proficiency Influences English Language Learners Performance on High Stakes Tests

PAPER TYPE: (Choose only 1) **SEM/YR COMPLETED:** _Spring 2011_

_____ Project
___x_ Thesis

SUPERVISOR'S SIGNATURE OF APPROVAL___________________________

Using key words or phrases, choose several ERIC descriptors (5 - 7 minimum) to describe the contents of your project. ERIC descriptors can be found online at:

http://www.eric.ed.gov/ERICWebPortal/Home.portal?_nfpb=true&_pageLabel=Thesaurus&_nfls=false

1. English (second language)	6. Mathematical Vocabulary
2. Mathematics	7.
3. High Stakes Tests	8.
4. Language Proficiency	9.
5. Accommodations for Testing	10.