USING PROC GLIMMIX TO ANALYZE THE ANIMAL WATCH, A WEB-BASED TUTORING SYSTEM FOR ALGEBRA READINESS

By

Otilia C. Barbu

_____________________

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF EDUCATIONAL PSYCHOLOGY
In Partial Fulfillment of the Requirements
For the Degree of
DOCTOR OF PHILOSOPHY
In the Graduate College

THE UNIVERSITY OF ARIZONA

2012
THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Otilia Catalina Barbu entitled Using PROC GLIMMIX to Analyze the Animal Watch, a Web-Based Tutoring System for Algebra Readiness Tutoring System and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Date: March 27, 2012
Dr. Deborah Levine- Donnerstein

Date: March 27, 2012
Dr. Ronald Marx

Date: March 27, 2012
Dr. Thomas Good

Date: March 27, 2012
Dr. Rebecca McGraw

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

________________________________________ Date: March 27, 2012
Dissertation Director: Dr. Deborah Levine- Donnerstein
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from the reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Otilia Catalina Barbu
ACKNOWLEDGEMENT

This dissertation could not have been completed without the help of many special people to whom I am deeply grateful.

First, I would like to express my most sincere thanks to the members of my committee: Dr. Deborah Levine-Donnerstein, Dr. Thomas Good, Dr. Ronald Marx, and Dr. Rebecca McGraw for their help, support, and patience. Without their continuous support, understanding, and willingness to give me guidance, I could not have completed this dissertation.

Second, I would like to thank Dr. Carol Beal for giving me the opportunity to work as a research assistant in her Animal Watch project, and the local school district for assisting me throughout the entire dissertation process. Their support was invaluable.

Finally, I must express my gratitude to my family for its continuous moral support, understanding, and encouragements.
TABLE OF CONTENTS

LIST OF TABLES .......................................................................................................................... 9

LIST OF FIGURES .......................................................................................................................... 10

ABSTRACT ...................................................................................................................................... 11

CHAPTER 1. INTRODUCTION ........................................................................................................ 12

Purpose and Contribution of the Study ....................................................................................... 14

Research Topics .......................................................................................................................... 14

CHAPTER 2. LITERATURE REVIEW ............................................................................................... 16

General Considerations about the Role of Bilingualism in Education ...................................... 16

Difficulties Met by ELs in Solving Algebra Word Problems ...................................................... 19

Classroom Influences on Mathematics Learning by ELs ............................................................ 20

ELs’ Performance on Mathematics Achievement Tests ............................................................. 21

Cognitively-Oriented Theoretical Framework ............................................................................. 22

Gender Related Studies in Mathematics Education .................................................................... 27

CHAPTER 3. METHODOLOGY ........................................................................................................ 35

Participants .................................................................................................................................. 35

Materials ..................................................................................................................................... 38

Procedure .................................................................................................................................... 41

Variables of Interest ................................................................................................................... 46

Data Cleaning ............................................................................................................................... 47

Introduction of the Model ............................................................................................................ 52

Summary of the Considered Models ............................................................................................ 60

Detailed Presentation of the Investigated Models ......................................................................... 61
TABLE OF CONTENTS - Continued

Model 1: Unconditional Means Model ................................................................. 61
Model 2: One Level-2 Predictor (ELL SECTION), no Level-1 Predictors .......... 63
Model 3: One Level-2 Predictor (SCHOOL), no Level-1 Predictors ................. 64
Model 4: One Level-1 Effect (FIRT TRY RATIO), no Level-2 Covariate ......... 65
Model 5: One Level-1 Effect (TOO HARD RATIO), no Level-2 Covariate ....... 66
Model 6: One Level-1 Effect (GENDER), no Level-2 Covariate ................. 67
Model 7: General Model with Level-1 Predictors (FIRST TRY RATIO, TOO
HARD RATIO, and GENDER), and level-2 Predictors (ELL SECTION and
SCHOOL) ................................................................................................................. 68
Model 8: The General Model with One Intersection Removed (TOO HARD RATIO
and SCHOOL) ......................................................................................................... 71
Model 9: The General Model with Two Intersections Removed (FIRST TRY
RATIO - ELL SECTION and TOO HARD RATIO - SCHOOL) ......................... 73
Model 10: The General Model with Three Interactions Removed (FIRST TRY
RATIO - ELL SECTION, TOO HARD RATIO - SCHOOL, and TOO HARD
RATIO - ELL SECTION)......................................................................................... 74
Model 11: General Model with all Interactions Removed............................... 76
Model Comparison............................................................................................... 77

CHAPTER 4. FINDINGS......................................................................................... 79
Descriptive Results ............................................................................................... 79
Model 1: Unconditional Means Model ............................................................. 81
Model 2: One Level-2 Predictor (ELL SECTION), no Level-1 Predictors ....... 84
TABLE OF CONTENTS - Continued

Model 3: One Level-2 Predictor (SCHOOL), no Level-1 Predictors ....................... 85
Model 4: One Level-1 Effect (FIRT TRY RATIO), no Level-2 Covariate .............. 86
Model 5: One Level-1 Effect (TOO HARD RATIO), no Level-2 Covariate ........... 87
Model 6: One Level-1 Effect (GENDER), no Level-2 Covariate ......................... 88
Model 7: General Model with Level-1 Predictors (FIRST TRY RATIO, TOO HARD RATIO, and GENDER), and level-2 predictors (ELL SECTION and SCHOOL) ............................................................................................................................................................................................. 89
Model 8: The General Model with One Interaction Removed (TOO HARD RATIO and SCHOOL) ................................................................................................................................................................................................. 92
Model 9: The General Model with Two Interactions Removed (FIRST TRY RATIO - ELL SECTION and TOO HARD RATIO - SCHOOL ) ............................................................................................................. 94
Model 10: The General Model with Three Interactions Removed (FIRST TRY RATIO - ELL SECTION, TOO HARD RATIO - SCHOOL, and TOO HARD RATIO - ELL SECTION) ............................................................................................................................................................................................................................. 96
Model 11: The General Model with all Interactions Removed ............................. 96

CHAPTER 5. CONCLUSIONS ..................................................................................... 104

Summary of the Study ................................................................................................ 104
Research Topic One .................................................................................................. 105
Research Topic Two .................................................................................................. 108
Research Topic Three ................................................................................................. 110
Limitations of the Study ............................................................................................. 112
Importance of the Study and Recommendations for Further Research ................. 113
APPENDIX A. CONSENT FORM FOR STUDENTS .................................................. 115
APPENDIX B. CONSENT FORM FOR PARENT/ GUARDIAN ......................... 117
APPENDIX C. SITE AUTHORIZATION LETTER ............................................... 121
APPENDIX D. WORD PROBLEMS ................................................................. 123
REFERENCES ................................................................................................. 125
LIST OF TABLES

Table 1. Demographic Characteristics, 2009-2010........................................................... 37
Table 2. Descriptive Statistics for Variables Included in GLMM Analyses ................. 81
Table 3. Model 1: Unconditional Means Model............................................................... 82
Table 4. Model 2: One Level-2 Predictor (ELL SECTION), no Level-1 Predictors....... 85
Table 5. Model 3: One Level-2 Predictor (SCHOOL), no Level-1 Predictors.............. 86
Table 6. Model 4: One Level-1 Effect (FIRST TRY RATIO), no Level-2 Covariate ...... 87
Table 7. Model 5: One Level-1 Effect (TOO HARD RATIO), no Level-2 Covariate.... 88
Table 8. Model 6: One Level-1 Effect (GENDER), no Level-2 Covariate ................. 89
Table 9. Model 7: General Model with Level-1 Predictors (FIRST TRY RATIO, TOO HARD RATIO, and GENDER), and Level-2 Predictors (ELL SECTION and SCHOOL) ................................................................................................................. 91
Table 10. Model 8: The General Model with One Intersection Removed (TOO HARD RATIO and SCHOOL) ................................................................................................................. 93
Table 11. Model 9: The General Model with Two Interactions Removed (FIRST TRY RATIO - ELL SECTION and TOO HARD RATIO - SCHOOL)............................ 95
Table 12. Model 11: The General Model with all Interactions Removed ................. 97
Table 13. Model Comparison ......................................................................................... 98
Table 14. Model 10: The Final Model.......................................................................... 100
LIST OF FIGURES

Figure 1. Cognitive Effects of Different Types of Bilingualism ........................................... 18

Figure 2. Bernoulli Distribution. Probability Density Function ........................................ 43

Figure 3. Q-Q plot for the Dependent Variable. ............................................................... 49

Figure 4: The logit function .............................................................................................. 54

Figure 5. First order regression with the covariate GENDER for the four classes analyzed .................................................................................................................................. 58

Figure 6. First order regression with the covariate FIRST TRY RATIO for the four classes analyzed .................................................................................................................................. 58

Figure 7. First order regression with the covariate TOO HARD RATIO for the four classes analyzed .................................................................................................................................. 59

Figure 8. The normality of residuals for the Model 10..................................................... 99
ABSTRACT

In this study, I investigated how proficiently seventh-grade students enrolled in two Southwestern schools solve algebra word problems. I analyzed various factors that could affect this proficiency and explored the differences between English Learners (ELs) and native English Primary students (EPs). I collected the data as part of the Animal Watch project, a computer-based initiative designed to improve the mathematical skills of children from grades 5-8 in the Southwest. A sample of 86 students (26 ELs and 60 EPs), clustered in four different classes, was used for this project. A Generalized Linear Mixed Model (GLMM) approach with the GLIMMIX procedure in SAS 9.3 showed that students from the classes that had a higher percentage of EL students performed better than those in the classes where the EL concentration was lower. Classes with more EL males were better at learning mathematics than classes with more EP females. The results also indicated: (a) a positive correlation between the students’ ability to solve algebra word problems on their first attempt and their success ratio in solving all problems, and (b) a negative correlation between the percentage of problems solved correctly and those considered too hard from the very beginning. I conclude my dissertation by making specific recommendations for further research.
CHAPTER 1
INTRODUCTION

The concept of equity in mathematics is certainly not new but, despite sustained
deper of the last four decades or more, it is still an issue in mathematics education
today. Language and gender inequities at all school levels are two of the most researched
areas.

Learning mathematics represents a complex school activity, which involves
acquiring a new vocabulary, manipulating new symbols and concepts in a specific
context, developing new thinking and reasoning skills, and communicating the results
and the steps of the deductive processes to the outside world. Some authors would even
consider that “the language of mathematics can be as challenging as a foreign language”
(Freeman & Crawford, 2008, p. 11). For English-language learners (ELs), the difficulties
associated with learning a specific mathematical vocabulary are amplified by an
incomplete knowledge of English.

It is essential to understand better the challenges facing ELs as they learn
mathematics because they consistently perform lower than English Primary students
is a major source of underachievement in schools” (Cuevas, 1984, p. 134). In addition,
he cites the landmark Supreme Court case of Lau v. Nichols (1974) which concluded that
“students who do not understand English are effectively foreclosed from any meaningful
education” (Cuevas, 1984, p. 134).
As to gender, the belief that women are not good at mathematics was prevalent throughout history, and different explanations were offered. In the beginning, it was argued their brains were too small (Henrion, 1997). Later, the study of mathematics by women was regarded as incompatible with their motherly attributes, or it was deemed a pure waste of time, since they lacked the talent or the necessary skills, compared with their male counterparts.

Over the last 40 years, many researchers have tried to understand why mathematics was considered a “male domain,” and many experimental studies have been carried out in different countries (US, Canada, UK, Australia, Singapore). As a result, researchers have proposed solutions to improve the situation, and differences between male and female roles in mathematics have narrowed.

To ensure equal mathematics education for all students, the National Council of Teachers of Mathematics (NCTM) published a comprehensive set of guidelines for teaching and learning mathematics at each grade level (NCTM, 2000). The central part of the Standards is represented by five educational principles: (a) a focus on equity and excellence for all students; (b) the necessity of developing a coherent curriculum; (c) the need to present students with teaching and learning experiences that builds on existing students’ knowledge base; (d) a careful assessment; and (e) a seamless technology implementation. Every state interpreted and implemented the NCTM recommendations independently; consequently, many contradictions and discrepancies exist. Moreover, the overall result of these efforts did not lead to the expected improvement in ELs mathematical performance: “Nationwide, 82% of Hispanic fourth-grade students are below proficient in mathematics (56% of whom are bellow basic), increasing to 88% of
Hispanic eight-grade students (50% of whom are below basic) “(Freeman & Crawford, 2008, p.11).

**Purpose and Contribution of the Study**

I conducted this study to investigate the possible differences in the mathematics achievement and behavior of middle school students, with respect to their English language abilities and gender. I collected the data as part of the web-based “Animal Watch” project, which was created to help middle school students’ learn mathematics. Using computers, the students solved algebra word problems involving various animals. The software provided help on request and monitored students’ activity.

**Research Topics**

The topics I planned to research and analyze are the following:

1. Since problems in Animal Watch are written in English, I expected ELs and EPs to perform differently. According to Cumming’s threshold hypothesis, ELs that have reached a specific level in English learning would perform better academically than EPs. Below this level, however, native English speaking students will achieve better results than ELs. Since each class in my study had a different ratio of EL to EP students, I expected variations in average mathematics scores among classes.

2. A considerable body of research has been dedicated to investigating the differences in mathematics performance between boys and girls. To date, no real
agreement exists in the scientific community. I planned to explore if boys and girls perform differently in a web-based testing environment.

3. I expected certain behavioral indicators could be used to predict the students’ success when solving algebra problems. For example, I expected students who could solve most problems on their first attempt would succeed better overall. Conversely, I expected students who considered the problems too difficult would not succeed.
CHAPTER 2
LITERATURE REVIEW

This chapter summarizes the main issues existing in mathematics education, focusing on language and gender. As the number of ELs in the United States has increased, so has the research aimed at creating optimal learning conditions for children with a primary language other than English. Similarly, gender equality has become a goal of modern societies. Men and women need equal education to compete in the marketplace. Furthermore, in a high-tech world, mathematics education has become even more important.

**General Considerations about the Role of Bilingualism in Education**

The role of bilingualism and its effects on the educational process has been examined for more than a century by the research community. Seen mostly as negative for a long time (Saer, 1923), views of bilingualism started to change in the 1960s and 1970s (Peal & Lambert, 1962; Skutnabb-Kangas & Toukomaa, 1976). Within the last four decades most of the studies pointed out the benefits of bilingualism in various areas of education and mental reasoning, and many researchers determined a slight superiority of some highly skilled bilingual students over their monolingual peers (Lambert, 1977; Cummins, 1978; Kessler & Quinn, 1987).

Concomitant with these observations, an increasing number of authors emphasized the fact that, in order to understand the role played by bilingualism in academic learning, bilingual students should not be categorized as an unidimensional
group, but rather differentiations should be made based on the students’ proficiency in the maternal (L1) and secondary (L2) languages (Cummins, 1976; Cummins, 1979; Clarkson, 1992). As an illustration of the role played by different levels of bilingualism, Cummins formulated the Threshold Hypothesis (Cummins, 1976; Toukomaa & Skutnabb-Kangas, 1977). Introduced mainly as a way of reconciling contradictory reports on the effects of bilingualism on learning performance, Cummins noticed that studies reporting negative effects on education were mainly performed on linguistic minorities, for which the maternal language (L1) was being replaced by the socially dominant one (L2), while studies reporting positive effects were performed on children whose skills in both the maternal language (L1) and the second language (L2) were well developed. According to Threshold Hypothesis, the level of linguistic competence attained by a bilingual child in first and second language may affect his or her cognitive growth in other domains. Cummins postulated the existence of two thresholds that would separate bilingualism levels and their cognitive effects in three regions (Figure 1). The lower threshold separates semilingualism (low level in both languages) with negative cognitive effects from dominant bilingualism (native-like level in one of the languages, but not in the other) with neither positive nor negative cognitive effects. The higher threshold separates the later domain from additive bilingualism (high levels in both languages) with positive cognitive effects.
The work of Barik and Swain (1976) presents evidence for the existence of the higher threshold. Their experiment, performed in Ottawa and Toronto, showed that high French (L2) achievers at grade 3 performed significantly better than low French (L1) achievers on two of the three Otis-Lennon IQ subtests. In addition, the IQ scores of high French achievers continued to grow over a three year period. Cummins (1979) formulated the Developmental Interdependence Hypothesis that states that the level of L2 competence that a bilingual child attains is influenced by the L1 competence level at the time when intensive exposure to L2 begins. If the L1 level is low, a “subtractive” bilingualism (Lambert, 1975) is expected to take place, in which the acquisition in the L2 domain decreases the L1 competence. On the other hand, for well developed L1 competences, the effect is “additive” (Lambert, 1975) in the sense that new acquisitions in either L1 or L2 have positive influences on the other language.
Early testing of Threshold Hypothesis in mathematics learning found that bilingualism is not necessarily a disadvantage (Dawe, 1983; Clarkson, 1992; Clarkson & Galbraith 1992; Secada, 1992). Dawe (1983) determined that the first language (L1) competence is an important factor for the ELs ability to reason in mathematics in English as a second language (L2) and that for both monolingual and bilingual students the knowledge of logical connectives in English is a crucial factor. The study determined that a higher English competence and better knowledge of the logical connectives correlated positively with better mathematics scores across all four linguistic minorities analyzed. Clarkson and Galbraith (1992) investigated the performance of Grade 6 students from five urban schools in Papua New Guinea and also determined that the linguistic competence in both first language (L1) and English (L2) determined the mathematics performance when the teaching language was English. More recent studies (Parvanehnezhad & Clarkson, 2008; Riordain& O’Donoghue, 2009) successfully replicated these results in Australia and Ireland, respectively.

One of the criticisms of the Threshold Hypothesis model is represented by the vagueness of the threshold levels (Baker, 2001; Riordain& O’Donoghue, 2009) and the need of an exact determination of the level of proficiency necessary in order to avoid negative effects and facilitate cognitive advantages.

**Difficulties Met by ELs in Solving Algebra Word Problems**

ELs’ mathematics performance is a multi-faceted issue, including challenges related to receiving instruction in a non-primary language and the relatively low rate of mathematics teachers who have received training in working with ELs (Coates, 2006). In
addition, research must consider the possible confounding of students’ language status with variations in socioeconomic status and other demographic factors. Multiple theoretical perspectives have been adopted in research on ELs, including sociocultural-oriented analyses of the classroom context and a more cognitively-oriented theoretical framework.

Classroom Influences on Mathematics Learning by ELs

Some researchers have analyzed how the classroom context experienced by ELs can impede or support mathematics learning. Cuevas (1984) noted that many ELs are confronted with the need not only to master a new language but also to learn the distinct features of academic discourse. Used in an instructional context, learning English is likely to be considerably more demanding than acquiring basic conversational proficiency. In the specific case of mathematics learning, students must come to understand the mathematics “register,” meaning the types of language used to convey mathematics concepts as well as the mathematics-specific meanings for words that may be familiar in other contexts (e.g., “table”) (Enyedy, Rubel, Castellon, Mukhopadhyay, Esmonde & Secada, 2008; Moschkovitch, 2002, 2005).

Other researchers have noted that mathematics instruction can sometimes rely on prior knowledge that may not always be shared by ELs. For example, a student who is not familiar with American sports may not understand a problem about football scores or baseball averages. These analyses make clear the need to ensure that teachers consider ELs’ experiences and take care to clarify forms of expression that may be confusing or ambiguous to a student who is learning English (Janzen, 2008). In addition, this
theoretical perspective holds that it is not appropriate for mathematics teachers to view ELs as being “deficient” but rather to identify the strengths of their prior knowledge and experiences that can be integrated into the examples used in the classroom (Freeman & Crawford, 2008; Rodriguez, 2009; Secada, 1996).

ELs’ Performance on Mathematics Achievement Tests

Other researchers have focused more specifically on investigating why ELs perform less well than EPs on mathematics achievement tests. Recent studies have demonstrated that reading proficiency in English is a significant predictor of mathematics test performance by ELs (Beal, Adams & Cohen, 2010; Guglielmi, 2008). In particular, the characteristics of mathematics test items that are especially challenging for ELs have been the subject of investigation (Echevarria, Powers & Short, 2006; Sireci & Khaliq, 2002; Solano-Flores, 2008). Martiniello (2008) identified a set of test items from a state mathematics achievement test that were known to be challenging for ELs, relative to EPs with similar overall proficiency in mathematics. She then worked with a small sample of ELs as they attempted the items. The results pointed to several item characteristics that impeded the ELs, including the use of low-frequency vocabulary terms (e.g., “spinner”) as well as words that were ambiguous in meaning, such as “one” (which could be a pronoun or reference to the number) and “change” (a difference, or money received in a financial transaction). Other test items referred to concepts and examples that would not necessarily be familiar to ELs from other cultures, such as coin-operated laundry equipment or a spelling bee competition.
Although the linguistic complexity of mathematics test items seems generally linked with poorer performance by ELs, the reason for this effect is not yet entirely clear. Some findings suggest that the nature of the difficulty experienced by ELs when solving mathematics problems is actually more general or diffuse than an occasional issue with unfamiliar vocabulary. For example, Martiniello (2008) found that the overall length of the word problem, rather than the inclusion of unfamiliar vocabulary, was most consistently associated with poor problem solving for ELs. Similarly, Wolf and Leon (2009) reported that the overall amount of academic vocabulary in word problem items was most predictive of item difficulty for ELs. A meta-analysis of studies looking at the impact of various vocabulary-related accommodations in testing, including bilingual dictionaries, glossary definitions, dual language tests, and simplified English found surprisingly little evidence that such word-level accommodations consistently led to better performance for ELs (Kieffer, Lesaux, Rivera & Francis, 2009).

Cognitively-Oriented Theoretical Framework

The activity of learning and knowledge formation are very complicated psychological processes, and they have been actively researched in cognitive psychology within the last three decades. Gutstein (1997) specifies that “a central idea is that knowledge is stored in the brain in richly intertwined semantic networks of ideas, concepts, facts and skills” (p. 711). The manner in which new knowledge is added to the already existing one is of crucial importance for understanding the process of mathematics learning in school. The new information becomes knowledge if it can be connected in a logical manner to the information already stored in the brain. At the
elementary level, the process of learning mathematics can be seen as made up of two fundamental blocks: a) learning new concepts (or definitions), and b) establishing new connections (proofs, relationships, theorems) between concepts. The mathematics knowledge cannot be understood or committed to long-term memory in isolated pieces; the new concepts must be related to existing ones in a logical manner, and based on a set of rules. Existing knowledge can be both formal and informal, and using it as a basis for further scientific accumulation is one of NCTM standards recommendations. Gutstein et al. (1997) mentions that “evidence confirms that helping teachers build on children’s informal knowledge in mathematics classrooms helps children use their intellect well, make meaning out of mathematical situations, learn mathematics with understanding, and connect their informal knowledge to school mathematics” (Gutstein et al., 1997, p.711).

Closely related to the idea of building on the children’s informal knowledge is the idea of testing that relies on the existence of this knowledge. Campbell et al. (2007) present the case of an EL pre-service elementary school teacher who had difficulties in solving a mathematics problem formulated as a baseball problem. The authors concluded that the teacher, in fact, possessed the necessary mathematics skills for solving the problem, but she did not possess the informal knowledge about the baseball game, knowledge taken for granted for a person born and raised in the US. This particular example illustrates the complexity of the EL situation; even if the students happen to understand the English words, in some cases they might not have a clear representation of their meaning. For example, the authors specify that the Spanish speakers might interpret the words “to do” as “to make”.
Starting from the presented case, the authors also point out another issue related to instruction that relies heavily on existing student knowledge: the limited amount of memory available for storing new information. Assuming that the teacher is aware of the limitations in informal knowledge of the ELs and that she is willing to compensate for it with additional explanations, these students will have to memorize an additional amount of information compared to their English speaking peers. By introducing this extraneous linguistic information, not necessarily connected to the mathematical concept under investigation, the ELs might run the risk of not being able to properly absorb the critical information. Under these circumstances, it is imperative that the teacher finds an informal common ground for all students before attempting to make use of this preexisting knowledge.

The issue of extraneous linguistic information overload is present not only in the teaching process, but also in the current standardized testing process. The authors mention that: “there is prima facie evidence that test writers are not linguistically or culturally aware of the difficulties that particular wording and phrasing in word problems cause students, especially those taking the tests in second or additional language.” (Campbell et al., 2007, p. 13)

The key to a correct understanding of the cognitive processes involved in teaching and learning mathematics is the association between cognitive load theory and the reflective abstraction processes. Cognitive load theory is centered on the attributes of memory, either long or short-term, while the reflective abstraction represents the process of active reenactment of the learned concepts, re-thinking of the problem and committing the remodeled structures to the long-term memory. The common liaison between these
two entities is represented by memory and we cannot understand the process of mathematics learning by analyzing them separately.

Without attention to issues of working memory, students are doomed to suffer inefficient and unproductive problem-solving techniques… Also, unless students have a stimulus to abstract by reflecting on the operations used in the solution of a problem, they risk never seeing a more inclusive picture. (Campbell et al., 2007, p. 15)

In the proposed new framework for integrating students’ culture and language in the teaching process, the authors suggest that “reflection on culture, language and socially situated prior experiences, in addition to reflection on mathematical content and students’ cognitive processes and understandings, be incorporated into models of mathematics teaching” (Campbell et al., 2007, p. 16). The teaching process should be continuously reevaluated and comprise multiple cycles of planning and instruction implementation. This framework is considered especially beneficial for the ELs: in the case when the teacher and the students are coming from different cultures, establishing a common ground requires continuous analysis and lesson planning.

In essence, the framework proposed by Campbell et al. (2007) for teacher education courses has four components: “(a) academic content; (b) mathematical and cognitive processes; (c) mathematical and contextual language; and (d) cultural/life experiences” (p. 20).

The academic content brings to attention the mathematics knowledge of the student – the richer this basis, the more the students will be capable to process and analyze the new information.
The second component, mathematical and cognitive processes, is concerned with identification and development of the cognitive processing skills needed for learning mathematics. Cognitive and metacognitive skills can be taught especially by using well-chosen examples, but also by questioning, planning, or drawing conclusions. The goal of the instruction, in the view of the authors “becomes one of enabling students to take control over their own learning through the practice and development of increasingly complex processes modeled by the teacher in activities and demonstrations, and in texts and materials” (Campbell et al., 2007, p. 22). In the case when students have difficulties in transferring strategies learned in one problem to another problem, the preferred method is the reduction of the goal specificity. According to this strategy, the students must be directed towards understanding the situation presented in the problem instead of focusing on the goal required by the problem. The authors conclude that “reducing the goal specificity in the early stages of teaching and learning a new principle reduces cognitive load” (Campbell et al., 2007, p. 22).

The third component of the proposed framework is represented by the relationship between mathematics and contextual language. In essence, the researchers are concerned with the degree in which the language used in the problems’ statement corresponds to the level of the English language of the ELs. One suggestion was for the teachers to enhance the role of natural language in instruction because it helps students mediate among mental processes, symbolic expressions, and logical organization, as well as “finding counterexamples and in developing arguments of validity” (Campbell et al., 2007, p. 23).

Cultural/life experiences are, in essence, concerned with the informal knowledge base that a student needs to possess in order to understand the mathematical concepts.
Since many of the ELs are coming from non-dominant cultural groups, their informal knowledge basis cannot be taken for granted. In the above example, not knowing details about the baseball game can prove a real impediment in understanding or applying simple mathematical rules.

**Gender Related Studies in Mathematics Education**

Elizabeth Fennema was one of the first researchers in the field of gender equity. In the mid-70s Fennema and Sherman proposed a widely accepted Mathematics Attitude Scales (MAS) system consisting of nine subscales, including mathematics as a Male Domain (MD). The purpose of this scale was to help categorize schools’ performance on mathematics learning between male and female students. In 2000 Fennema summarized her work on achieving gender equity in mathematics:

1. Gender differences in mathematics may be decreasing.
2. Gender differences in mathematics still exist in
   - Learning of complex mathematics
   - Personal belief in mathematics
   - Choice that involves mathematics
3. Gender differences in mathematics vary by
   - Socioeconomic status and ethnicity
   - School
   - Teacher
4. Teachers tend to structure their classroom in favor of male learning
5. Interventions can achieve equity in mathematics.
The issue of gender in mathematics is extremely complex. Indeed, large variations between groups of females exist. Many women succeed and make careers in mathematics; however, the number of females teaching mathematics at the college level is considerably smaller than the number of males. Furthermore, attitudes about gender among teachers and schools vary widely, often reflected in complicated mathematical tasks assigned across grade levels.

The existing experimental results are categorized in Fennema’s study from two different perspectives: cognitive science and feminism.

From the cognitive science point of view, the experiments could not reveal any differences in how the boys and girls solve arithmetic problems, and the study infers that the processes used by both genders to make sense of arithmetic are essentially the same. On the other hand, while the teachers believe that the attributes of boys and girls who succeeded in mathematics were similar their knowledge about which boys were successful was more accurate than their knowledge about which girls were successful. Also, teachers reported that they thought more about boys than about girls during instruction, but they described both genders’ performance in similar terms.

Other researchers investigated the possible relationship between confidence in learning mathematics and the actual learning of mathematics. It has long been assumed that lower confidence contributes to gender differences in mathematics. Fennema, however, could not find any decisive evidence that lower confidence levels in girls are responsible for lower performance. In contrast, she did find experimental evidence that when the teachers make instructional decisions based on their knowledge of individual children, overall gender differences disappear.
From the feminist perspective, scholars have argued that most of our beliefs, perceptions, and scientific methodologies are dominated by male perspectives (Fennema, 2000). Some scholars even tried to define what a feminist approach to the study of mathematics education might be. Their approach to the problem of gendered mathematics is not to look at the subject, but to examine the way that people think and learn within the subject. Fennema (2000) suggested that girls may learn mathematics better in cooperative settings than in competitive settings, which favor boys. She pointed out that the existing research provides a rich set of results about different variables affecting gender differences in mathematics, and these results decrease the existing gap between boys and girls. She further suggested that future research would need to focus on new questions related to equity in mathematics, and new methodologies should be considered in evaluating them.

In a comprehensive study, Seegers and Boekaerts (1996) investigated gender differences related to mathematics achievement. The researchers collected data from 186 students from nine schools in the Netherlands, using questionnaires and tests. Boys and girls were evenly represented. The study measured the students’ mathematics performance, as well as the goal orientation, attributions, self-concept of ability, and task-specific appraisals. The research investigated the relationship between mathematics learning, trait-like variables (self-concept of ability, goal orientation, and attributions), and task-specific appraisals (personal relevance, task attraction, and subjective competence) that relate to mathematics learning and to the differences that exist between boys and girls. According to the authors, the group of cognitive variables called "self-referenced cognitions" (Seegers & Boekaerts, 1996)—variables which describe the
perception the students have about themselves, about their abilities, feelings, and skills—was correlated with the gender differences in mathematics learning.

The first of the trait-like variables, the academic self-concept of ability, is considered to have a strong influence over the ability to perform specific tasks and can directly influence the outcome of the learning process. The academic self-concept of ability is critical in motivating students for work, and different studies found differences between boys and girls in the way they estimate their abilities in mathematics learning. Boys constantly showed a higher level of confidence in their aptitudes, even when no differences in mathematics achievements were present.

The second trait-like variable considered in the study is the goal orientation. This variable is related to mathematics achievements, and it is age dependant. Younger students associate success with effort while older students associate success with mathematical abilities.

The attributions variable is related to outcome expectation as well as to the relationship made by the students between the learning activity and success. The researchers pointed out that, in general, success is associated with internal factors while failure is linked with external conditions. Differences between genders exist. Boys consider that their success is due to their abilities in a higher degree than girls. On the other hand, girls believe more often than boys that their failure is due to their lack of skills.

The task-specific appraisals considered in this study were personal relevance, task attraction and subjective competence. According to the authors, gender differences in all task-specific appraisals have been reported before. The previous research found that
boys have a higher level of self-efficacy, while girls are more willing to learn. Moreover, the personal relevance represents a measure of the value associated with a task. In addition, the subjective competence is made up of three components: “self-efficacy, success expectation and perceived level of difficulty” (Seeger & Boekaerts, 1996, p. 220).

The authors showed that boys outperformed girls in mathematical abilities, and their capacity for learning mathematics was superior to that of girls’ (the self-concept of ability). On the other hand, girls were more eager than boys to invest hard work in studying mathematics. The authors also found gender-related differences for the other two trait-like variables: goal orientation and attributions. Girls described more often than boys that their lack of success was due to their lack of abilities while boys were more goal oriented. The results of the study suggested that boys’ learning is more favorable in a competitive climate, while girls would do better in a cooperative setting.

Hyde et al. (1990) performed a meta-analysis of 100 studies concerning gender differences in mathematics. They noted that gender differences in mathematics performance have always existed. These differences are not apparent in early childhood, but they appear in adolescence. Overall, high school girls did not perform quite as well as boys when solving problems. The authors argued that this discrepancy requires further research. Usually, boys do better in tasks involving high cognitive complexity (problem solving), and girls do better in tasks involving less cognitive complexity (computation).

The theoretical models for gender performance assume that males outperform females but, in the authors’ opinion, this view is in need of reassessment, using modern tools of meta-analysis. Their meta-analysis was based only on published studies reporting psychometrically developed mathematics tests, and large-sample normative
data were obtained for the following tests: American College Test (ACT); Graduate Management Admission Test (GMAT); Scholastic Aptitude Test-Quantitative (SAT-Q); SAT Mathematics 1 and 2; Dental Admission Test (DAT); Graduate Record Examination-Quantitative (GRE-Q); and GRE-Mathematics. One hundred usable data sources, yielding 259 independent effect sizes were included in this meta-analysis. The effect size computed was $d$, defined as the mean for males minus the mean for females divided by the mean within-sexes standard deviation. Positive values of $d$ represent superior male performance and negative values represent superior female performance.

The mean magnitude of the gender difference averaged over all studies was .20. When the SAT data were excluded, the value of $d$ became .15. The researchers found that males performed better than females, but the difference was small. The results also show that the gender difference has declined over the past three decades. The value of $d$ was .31 for the studies published in 1973 and earlier and .14 for studies published in 1974 and after.

In addition, girls slightly outperformed boys in elementary and middle school, but a moderate male superiority appeared in high-school ($d= .31$) and continued in college ($d= .41$). The age trends were a function of the cognitive level investigated. Females were superior in computation in elementary and middle school, but no difference existed in high-school or after. The most important finding was that the problem solving abilities changed with age from middle school, when the girls had a slight advantage over boys, to high school and college, when the boys performed better.

The results of this meta-analysis study provided little support for male superiority in mathematics. Females were superior in computation, and no gender differences in
understandings of mathematics were found. Males tended to outperform females in problem-solving tasks beginning with high school. Moreover, given this small difference in mathematical abilities, the reason for female under representation in college-level mathematics courses and in mathematics-related occupations may be due to other internal factors (e.g., internalized belief systems about mathematics) or external factors (e.g., sex discrimination in education and employment).

More recent studies, (McGraw et al., 2006; Amelink, 2009; Ai, 2002; Lubienski et al. 2004; Lubienski & Bowen, 2000) continued to show relatively small differences of gender achievements in mathematics. McGraw et al. (2006) found a small but significant difference in favor of boys. She analyzed the results published by the U.S. National Assessment of Educational Progress (NAEP) from 1990 to 2003. This difference stayed relatively constant over the investigated period and its absolute magnitude was influenced by race, social economic status (SES), and the mathematical area under consideration. In all situations, the gender difference between the highest performers was larger than the difference between lower performers when examined at the same percentile levels. Amelink (2009) also determined that overall math performance for both genders has increased during the last two decades, with boys keeping, in general, a slight advantage. Her results were based on investigations of various national evaluations: American Association of University Women (AAUW), 1992; National Association of Educational Progress (NAEP), 2007a; National Educational Longitudinal Survey (NELS), 2004; and National Science Foundation (NSF), 2008). However, Grade Point Average (GPA) results showed that girls surpassed boys both in overall scores and in mathematics scores for the entire period investigated, from 1990 to 2005. By citing US and international
research in the field, Amelink (2009) establishes that classroom climate contributes significantly to the overall results. Countries with more equal gender expectations, compared to the US, exhibit a smaller or non-existent gap in mathematics achievement between genders (Guiso et al., 2008). Mullis et al. (2000) found that girls generally outperform boys in mathematics abilities at early age, but this advantage vanishes by the age of 17. These authors also found that girls perform better in reading graphs, computation, and algorithmic problem solving.

Different expectations play a major role in gender performance gaps because they erode girls’ self-confidence and perceived efficacy. Stipek and Granlinski (1991), found evidence that girls' achievement-related beliefs were generally lower than boys' achievement-related beliefs at both elementary and high-school levels. Yet, these differences were found to disappear once both genders received the same positive feedback for their aptitudes and performance. Brophy and Good (1970) found that boys have more interactions with teachers than girls, with teachers directing most of their evaluative comments (both positive and negative) towards boys. In addition, de Boer et al. (2010) determined a relationship between teacher expectation and students’ characteristics, including gender. Moreover, the authors found that this expectation bias presented a long-term effect on students’ performance.
In this chapter I describe the methodology used to investigate the rate of success of ELs and EPs on mathematics word problem solving. I provide information about the research context, as well as the methods I used to collect and analyze the data.

Participants

Eighty-six seventh-grade students (26 ELs and 60 EPs) attending two public middle schools—LCMS and AMS—from the same school district participated in this study. The district classified the ELs by using four language proficiency tests: (1) Arizona English Language Learner Assessment (AZELLA), (2) the new Arizona English Language Learner Assessment AZELLA 2 Form AZ-2 (AZ2), (3) Stanford English Language Proficiency (SELP), and (4) IDEA Oral Language Proficiency Test (IPT). From a total of 26 ELs the district classified four as having an intermediate English status and the other 22 as having a proficient English status. No students with a low English status were identified. In addition, following Pray’s suggestion, I used social economic status (SES) and enrollment in special education to classify the ELs native language skills (Pray, 2005). I used students’ enrollment in the lunch program to determine SES. Those students who received a free or reduced lunch were considered “middle level in native language” and those who paid full price for lunches were considered “upper level in native language.” Those students who were enrolled in a special education program were considered “low level in native language.” As a result, of 26 ELs, two students were classified as having a low level in their native language, four students as having a middle
level in their native language, and the rest as having an upper level in their native language.

Within the district, middle school students were diverse. Males made up 51% of the population. The overall student population was 52.4% white, 36.5% Hispanic, 4.6% African American, 4.5% Asian American, and 2% Native American. The schools where I conducted my study were diverse as well (see Table 1). In both schools the percentage of male students made up 51% of the population with a slight variation of less than 2% by grade level. The number of Asian American (3.20% to 3.90%) and Native American (3.60% to 4.80%) students was fairly stable at all campuses. The African American population varied much more dramatically, from 8.80% at AMS to 2.40% at LCMS. The white population ranged from 21.30% at AMS to 24.40% at LCMS. The Hispanic population ranged from 61.90% at AMS to 65.80% at LCMS. At AMS, 7th graders comprised 35.70% of the total number of students; at LCMS, they comprised 31.60%.
Table 1

*Demographic Characteristics of the sample and district (%), 2009-2010*

<table>
<thead>
<tr>
<th></th>
<th>AMS n=711</th>
<th>LCMS n=619</th>
<th>Middle School District n=3435</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian American</td>
<td>3.20</td>
<td>3.90</td>
<td>4.50</td>
</tr>
<tr>
<td>African American</td>
<td>8.80</td>
<td>2.40</td>
<td>4.60</td>
</tr>
<tr>
<td>Hispanic</td>
<td>61.90</td>
<td>65.80</td>
<td>36.50</td>
</tr>
<tr>
<td>Native American</td>
<td>4.80</td>
<td>3.60</td>
<td>2.00</td>
</tr>
<tr>
<td>White</td>
<td>21.30</td>
<td>24.40</td>
<td>52.40</td>
</tr>
<tr>
<td>English language learners</td>
<td>2.60</td>
<td>1.70</td>
<td>0.90</td>
</tr>
<tr>
<td>7th grade distribution</td>
<td>35.70</td>
<td>31.60</td>
<td>7.30</td>
</tr>
<tr>
<td>Eligibility for Free/Reduced Lunches</td>
<td>88.00</td>
<td>78.00</td>
<td>42.50</td>
</tr>
<tr>
<td>Distribution of Special Education Students</td>
<td>16.60</td>
<td>15.20</td>
<td>13.40</td>
</tr>
<tr>
<td>Distribution of REACH Students</td>
<td>10.10</td>
<td>14.80</td>
<td>14.20</td>
</tr>
</tbody>
</table>

Source: *Data Book 2009-2010* [Data file].

The students participated in this project as part of their mathematics classroom instruction. I obtained informed consents from parents or legal guardians as well as all student participants.

The fact that the number of ELs and EPs is unbalanced was not a problem in this study. Multi level modeling is appropriate for analyses of unequal sample sizes.

Unequal sample sizes at each of the levels pose no problems and are, indeed, expected
Group sizes may be as small as one, as long as other groups are larger (Snijders & Bosker, 1999).

**Materials**

In this study, middle school students worked with an existing web-based tutoring system for mathematics word problem solving called Animal Watch. Students worked with the software in their classrooms or school computer lab under the direction of their teacher during their mathematics class. The computer program connects mathematics to science and presents the students with different mathematics problems centered on various animals. After login, the student starts at the Dashboard. The Animal Watch system keeps track of a student’s place in the problem sequence. The Dashboard also contains a message of the day that can be personalized (e.g., “Joel, great work so far! If you finish this set early, go to Skill Builder”).

Problem sets are introduced with scientific information about endangered species and indicate the type of mathematics problems. Problem pages have a standard layout. If a student submits an incorrect response, his/her input is identified and problem-specific feedback is provided. Each problem has two distinct problem-specific feedbacks (short answer and long answer), and multiple help items that can be accessed at any time. The software can provide help, upon request, in the form of additional explanations and demonstrations and records the student’s activity for future analysis.

The Help menu includes a completed example, an interactive example with a visual representation of the mathematics concept, and a video lesson. Once the problem loads on the screen, the student can take various actions: he/she could enter the
correct/incorrect answer up to three attempts, skip the problem, or click the "Help" icon multiple times. The word problems vary in text characteristics (e.g., grade-level readability, number of words, and inclusion of scientific vocabulary) as well as in the mathematics operation required to solve the problem (e.g., arithmetic, fractions, or simple algebra).

The program has three levels of difficulty: Introductory, Intermediate, and Challenging. Within the Introductory level the program contains Easy-Math and Easy-English problems (EM-EE). Within the Intermediate level the program contains Easy-Math and Hard-English (EM-HE) and Hard-Math and Easy-English (HM-EE) problems. Within the Challenging level the program contains Hard-Math and Hard-English (HM-HE) programs. (For examples of each type of problem, see appendix A.)

In this study I examined different students solving different problems with the same Intermediate level of difficulty. Easy mathematics problems were considered those that required only one single-digit addition; hard mathematics problems contained a combination of multi-digit multiplications and additions. The language difficulty was established by using the REAP readability assessment software developed at Carnegie Mellon University (http://reap.cs.cmu.edu). Many of the mathematics problems were written in both easy-language and hard-language. The increased language difficulty was obtained by changing the vocabulary and the syntax while keeping the word count constant. On the difficulty scale provided by REAP software (from 1 to 10), the easy-language problems were rated at level 3, while the difficult-language problems were rated at level 8.
After the student sees a specific problem, he/she has several options. He/she may solve the problem or rate the problem as too hard. Or he/she may attempt to solve the problem up to three times, rating it too hard at any time during the process. After completing a problem, a student may return to the Dashboard, or continue to the next problem. He/she may also view a progress report at the completion of a problem set, a skill builder, or an entire lesson objective.

Skill Builders are sets of ten true or false statements designed to give quick and frequent practice with specific mathematics skills. The program provides immediate feedback about accuracy. Students choose among mathematics facts and vocabulary skill builders that support the current learning objective. The goal of the Skill Builders is to improve student accuracy and speed when solving word problems. Upon completion, a screen provides details on performance and average time per problem. From here, students can return to the Dashboard to view a more detailed report or continue with Skill Builders and Problem Sets.

The program provides teachers with Class Flash Reports and detailed Student Score Reports. The Class Flash Reports contain the following: (1) percent of problems solved correctly by ELs and EPs and class averages (Correct); (2) percent of students who declined to answer a problem (Give up); (3) percent of students who didn’t answer correctly in three attempts (Strike out); and (4) percent of students who used help for one or more problems (Help). From the Class Flash Reports each student name links to an additional report, called Student Score Report. The Student Score Report shows how a student performed on each problem and for the entire learning objective, as well as how long he/she spent on each problem, learning objective, and session.
Procedure

To complete my study, I needed to conduct statistical analyses. In general, statistical models are mathematical representations of population behavior. When a particular statistical model is used to analyze a particular set of data, “you implicitly declare that this population model gave rise to these sample data. Statistical models are not statements about sample behavior; they are statements about the population process that generated the data” (Singer & Willett, 2003, p.46). Statistical models are expressed using parameters (intercepts, slopes, variances, and so on) that describe specific characteristics about a population.

For this study, I employed a Generalized Linear Mixed Model (GLMM) using the GLIMMIX procedure in SAS for analysis. Generalized Linear Mixed Models (Breslow & Clayton, 1993), also known as hierarchical generalized linear models (HGLMs) or generalized linear models with random effects (Schall, 1991) are analyses for multilevel data with nonlinear structural models and nonnormally distributed errors. Leonard (1972) first discussed Bayesian hierarchical models for binomial data; Zellner and Rossi (1984) gave an overview of Bayesian methods for binomial regression models; and Johnson and Geisser (1985) introduced general Bayesian predictive and estimative case deletion diagnostics that apply to binomial regression. While the Hierarchical Linear Models (HLM) deal with continuous normally distributed data, the GLMM is appropriate for data with correlations or nonconstant variability where the response is not necessarily normally distributed and having any distribution in the exponential family: (1) discrete (binary, binomial, Poisson, and negative binomial distributions), or (2) continuous (normal, beta, gamma, and chi-square distributions).
In my study each student had a set of problems to solve with only two (binary) outcome values: success (1) or failure (0). Responses with only two possible outcomes are called Bernoulli random variables, and observations on them are called Bernoulli trials. The goal of the studies that observe the Bernoulli responses is to estimate the probability of a particular outcome and see how treatments or other predictor variables affect the probability. Probabilities are estimated by sample proportions, that is, the number of times the outcome of interest occurs divided by the number of independent Bernoulli trials observed. Observations of this type have a Bernoulli distribution (Figure 2), and it is one of the best known non-normal probability distributions.
Figure 2. Bernoulli Distribution. Probability Density Function (PDF), \( f(0) = 1 - p \), \( f(1) = p \) varies with the outcome of the single trial.

While the standard linear models methods assume normality and homogeneity of variance, Bernoulli response variables violate these assumptions. In general, violations of normality compromise the estimation of coefficients, and the calculation of confidence intervals. If a distribution is significantly non-normal, the confidence intervals may be too wide or too narrow. In this case, the results may not be trustworthy, resulting in Type I or Type II errors. Traditionally, the main tool to adapt non-normal data to regression models has been transformations. More recently, generalized linear models have provided a systematic approach to adapting linear model methods.

There are several problems with transformations. First, most commonly used transformations are intended to stabilize the variance, but they fail to address the problem of skewness. Second, transformations express the data on scales that are unfamiliar to
those who use the results of the analysis, creating problems of interpretations (Little et al., 2006). Conventional generalized linear model programs such as GENMOD do not provide for random effects, and mixed models such as PROC MIXED do not provide for non-normal errors and thus are not appropriate for use with dependent variables that do not present a normal distribution.

The GLIMMIX procedure is an extension of these two procedures and allows for the response to have a non-Gaussian distribution with arbitrary skewness and kurtosis and provides for random effects (Little et al., 2006). In addition, GLIMMIX allows two different outcome variables, each with a different distribution, to be modeled jointly. GLIMMIX fits generalized linear mixed models based on linearization. The default estimation method uses restricted pseudo-likelihood estimation with an expansion around the current estimate of the best linear unbiased predictors of the random effects. The technique is a double iterative process. Nonlinearity is removed by using a first-order Taylor series. The generalized linear mixed model is approximated based on current values for the covariance parameter estimates. When convergence is obtained, new parameter estimates are used to update the linearization, which results in a new linear mixed model. The process is repeated with the current pseudo data. The procedure continues until parameter estimates between successive linear mixed model fits are within a specified amount.

Generalized linear mixed-model programs such as the GLIMMIX macro or, on some cases, PROC NL MIXED, are better suited for such models (Little et al., 2006). While the NLMIXED procedure for nonlinear mixed models was added in SAS 8, the GLIMMIX procedure for generalized linear mixed models was added on SAS 9.1 in
2006, and it is considered the best procedure for Bernoulli responses (Little et al., 2006).

In addition, the GLIMMIX method can simultaneously take into consideration the non-normality of the distributions and the correlations present in multiple outcomes without sacrificing the power to detect true differences (Little et al., 2006).

For HGLM analyses that use nonlinear link functions, a distinction can be made between unit-specific results and population-average results. The unit-specific model describes processes captured by level-1 coefficients that differ over the population of the level-2 units. The population-average models, in contrast, focus on group-level variables rather than the varying effects of individual-level covariates, and give answers to population-average questions. As Raudenbush and Bryk (2002) pointed out, “the population-average results can be deduced as one characteristic of the distribution of the unit-specific results” (p.304). Thus, the unit-specific models are richer and better suited to my analysis.

According to Singer (1998), multilevel models can be written in at least three different ways: (a) by expressing separate equations at each level; (b) by writing separate equations at each level and then substituting in the level-1 equation to arrive at a single equation; and (c) by writing a single equation that specifies the multiple sources of variation. In this study I wrote the equations at levels 1 and 2, and then I substituted into the level-1 equation to arrive at a single equation representation (the combined equation). The combined equation has two parts: (a) the structural part or fixed effects, and (b) the stochastic part or random effects. This distinction parallels the classical psychometric difference between true scores and measurement error.

In my analyses, I used the notation from Bryk and Raudenbush (1992).
Variables of Interest

In this study the students were clustered in four different classes in two schools from the same district in the Southwest. The concentration of ELs in each class was different. For the purpose of this analysis, the students and all explanatory variables directly related to them were at Level 1, while all explanatory variables directly related to the four clustering classes were at Level 2. The analysis of variables was completed in three steps. The first step consisted in testing the significance of variables in SAS9.3. From a total of 12 possible variables (SECTION, ELL SECTION, GENDER, SCHOOL, FIRST TRY, SECOND TRY, THIRD TRY, NUMBER ATTEMPTS, STRIKEOUT, TOO HARD, NUMBERS OF PROBLEMS SOLVED CORRECTLY, USED HELP) only six, including the outcome, were significant and used in my model.

The second step involved checking for intercorrelations between the predictor variables in order to avoid multicollinearity. In SAS, this is usually done by using the REG procedure and examining the values of variance inflation factor (vif), tolerance (tol) and collinearity (collin) terms. The REG procedure is a general-purpose procedure for linear regression that handles multiple regression models and produces collinearity diagnostics. The value of tolerance represents the proportion of variance in a given predictor that is not explained by all the other predictors, while the value of variance inflation parameter is calculated as the inverse of tolerance. A good rule of thumb is to make sure that the condition index provided in the Collinearity Diagnostics section is less than 30 (Neter et al., 1989).
In the third step I transformed the significant variables into proportions of favorable responses because different students solved different problems with the same grade of difficulty. Thus, the variables of interest in this project are:

1. SUCCESS RATIO: the ratio of problems solved correctly by each student over the total number of problems presented (outcome variable or dependent variable).

2. FIRST TRY RATIO: the ratio of problems solved correctly from the first attempt over the total number of problems solved correctly (Level-1 explanatory variable).

3. TOO HARD RATIO: the ratio of problems considered by each student to be too hard over the total number of problems presented (Level-1 explanatory variable).

4. GENDER: a dummy variable (Level-1 explanatory variable).

5. ELL SECTION: the ratio of ELs (the number of ELs over the entire number of students in each of the four classes/sections) (Level-2 explanatory variable).

6. SCHOOL: a dummy variable (Level-2 explanatory variable). [Note, with only two schools, this variable is considered a fixed predictor at the second level rather than specifying an additional third level of analysis.]

**Data Cleaning**

I found no missing data on preliminary inspection. I found by visual inspection two outliers in the outcome variable of the Q-Q plot and eliminated them. I made this decision based on Tabachnick and Fidell’s (2007) suggestion to delete the outliers that were not members of the population from which I intended to sample. In general, outliers are found in both univariate and multivariate situations, and they lead to both
Type I and Type II errors, frequently with no clue as to which effect they have in data analysis. In my case, potential explanations for these outliers could be the lack of interest of the students in solving mathematics word problems or multiple guessing.

I estimated the normality of the data by three methods: (a) visual inspection of the Q-Q plot; (b) calculation of skewness and kurtosis of the histogram; and (c) use of the Kolomogorov-Smirnov test for normality (a value of p > .05 is desired).

In the most general definition, a Q-Q plot is used to determine if a sample of data follows a specific statistical distribution (usually normal), and it is obtained by plotting the sample quantiles versus the quantiles of the assumed distribution. If the sample was extracted from the assumed distribution, the plot will be a straight line. Examination of the Q-Q plot revealed a deviation from the straight line from lower left to upper right, indicating a departure of the normality for the dependent variable.
Two components of normality are skewness and kurtosis. When a distribution is normal, these values are zero, indicating that data are perfectly symmetrical. Bulmer (1979) suggested the following rule of thumb for sample skewness: (a) if skewness is less than $-1$ or greater than $+1$, the sample distribution is highly skewed, (b) if skewness is between $-1$ and $-\frac{1}{2}$ or between $+\frac{1}{2}$ and $+1$, the sample distribution is moderately skewed, and (c) if skewness is between $-\frac{1}{2}$ and $+\frac{1}{2}$, the sample distribution is approximately symmetric. With a skewness of $-0.86$, calculated in SPSS, I concluded that the sample data were moderately negatively skewed.

The present data set is just one sample drawn from a population, and the sample could be skewed even though the population is symmetric. In this case, I needed to inspect the skewness for population. According to Cramer (1977), dividing the sample skewness ($G_1$) by the standard error of skewness (SES) provides the test statistic ($Z_{g1}$), which measures how many standard errors separate the sample skewness from zero:
\[ Z_g1 = \frac{G1}{\text{SES}} \]

where

\[ \text{SES} = \left\{ \frac{6n(n-1)}{(n-2)(n+1)(n+3)} \right\}^{1/2} \]

The same author suggested the following rule of thumb for population skewness:
(a) if \( Z_g1 < -2 \), the population is very likely skewed negatively; (b) if \( Z_g1 \) is between \(-2\) and \(+2\), the conclusion about the skewness of the population can not be reached (it might be symmetric, or it might be skewed in either direction); and (c) if \( Z_g1 > 2 \), the population is very likely skewed positively. In my case, the standard error of skewness is the following:

\[ \text{SES} = \left\{ \frac{6n(n-1)}{(n-2)(n+1)(n+3)} \right\}^{1/2} = \left\{ \frac{(6 \times 84 \times 83)}{(82 \times 85 \times 87)} \right\}^{1/2} = .26 \]

with the test statistic:

\[ Z_g1 = \frac{-0.86}{.26} = -3.3 \]

There is a debate in the literature about the relationship between the sample and population skewness. Joanes and Gill (1998) pointed out that sample skewness is an unbiased estimator of population skewness for normal distributions, but not others. For this reason, I computed the 95% confidence interval of population skewness according to the formula:

\[ 95\% \text{ confidence interval of population skewness} = G1 \pm 2 \text{ SES}, \]

and the further the sample skewness is from zero, the more skeptical we should be. In my case, the 95% confidence interval of population skewness was

\[ G1 \pm 2 \text{ SES} = -0.86 \pm 2 \times .26 = -.86 \pm .52 = -1.38 \text{ to } -0.34. \]
With a test statistic of $Zg1 = -3.3$ which was not included in our 95% confidence interval, I concluded that the population from which I sampled the present data was also negatively skewed.

Kurtosis has to do with the peakedness of a distribution, and the reference standard is a normal distribution, which has a kurtosis of 3. The statistical packages subtract 3 before reporting kurtosis so that the expected value is zero. The sample kurtosis calculated in SPSS had a value of .61 indicating a slight deviation from normality.

To determine the test statistic ($Zg2$) for population kurtosis, which tells how many standard errors the sample kurtosis is from zero (Cramer, 1977), I divided the sample kurtosis ($G2$) by the standard error of kurtosis ($SEK$),

$$Zg2 = G2/SEK$$

where

$$SEK= 2SES \left[\frac{(n^2-1)}{(n-3)(n+5)}\right]^{1/2}$$

Cramer (1977) suggested the following rule of thumb for population kurtosis: (a) if $Zg2 > +2$, the population very likely has positive kurtosis; (b) if $Zg2 < -2$, the population very likely has negative kurtosis; and (c) if $Zg2$ is between $-2$ and $+2$, any conclusion about the kurtosis cannot be reached.

In my case, the standard error of kurtosis was

$$SEK=2SES \left[\frac{(n^2-1)}{(n-3)(n+5)}\right]^{1/2} = 2 \times .26 \times \left[\frac{(84^2-1)}{81 \times 89}\right]^{1/2} = .51$$

with the test statistic

$$Zg2 = G2 / SEK = .61 / .51 = 1.19,$$

concluding that any decision regarding the population kurtosis cannot be obtained.
Finally, I tested the normality of the data with the Kolmogorov-Smirnov test in SPSS, and the p-value was $p = .0023$ indicating a departure from normality. A value of $p = .50$ or higher for this test is usually considered to describe a population with zero skewness.

**Introduction of the Model**

To analyze the data, I needed to introduce general considerations about a two-level model. In a multilevel logistic regression, the model constructed for the outcome is linear based on log odds (the natural logarithm of the odds) or logit. The level-1 model in GLMM consists of three parts: (1) a sampling model; (2) a link function; and (3) a structural model.

The sampling model for a two-level GLMM has the level-1 outcome $Y_{ij}$ associated with the $i^{th}$ student in the $j^{th}$ class defined as the number of “successes” in $m_{ij}$ trials with the $\phi_{ij}$ probability of success on each trial. I write

$$Y_{ij} \mid \phi_{ij} \sim B(m_{ij}, \phi_{ij})$$

(1.0)

to denote that $Y_{ij}$ has a binomial distribution with $m_{ij}$ trials and the probability of success on each trial as $\phi_{ij}$. Equation (1.0) uses two subscripts, $i$ and $j$, to identify individuals and classrooms, respectively. For the present data, $i$ runs from 1 through 86 (for the 86 students), and $j$ runs from 1 through 4 (for the 4 classrooms). According to the binomial distribution and in agreement with Raudenbush and Bryk (2002) notation, the expected value and the variance of $Y_{ij}$ are

$$\text{E}(Y_{ij} \mid \phi_{ij}) = m_{ij} \phi_{ij},$$

(1.1)

$$\text{Var}(Y_{ij} \mid \phi_{ij}) = m_{ij} \phi_{ij} (1 - \phi_{ij})$$

(1.2)
When \( m_{ij}=1 \), the outcome \( Y_{ij} \) is a binary variable taking as values zero or unity.

The link function used for binary and binomial outcomes is the logit link that is the log of the odds of success

\[
\eta_{ij} = \log \left( \frac{\varphi_{ij}}{1 - \varphi_{ij}} \right)
\]

where the *odds* are defined as the probability of success to the probability of failure. If the probability of success is .5 (\( \varphi_{ij}=.5 \)), the probability of failure is .5 (1-\( \varphi_{ij}=.5 \)), the odds of success are \( \varphi_{ij} / (1 - \varphi_{ij}) = .5 / .5 = 1 \) and the log odds or “logit” is \( \log(1) = 0 \). When the probability of success is less than .5, (\( \varphi_{ij} < .5 \)), the probability of failure is greater than .5 (1-\( \varphi_{ij} > .5 \)), the odds of success are less than 1 and the logit is negative (\( \eta_{ij} < 0 \)). When the probability of success is greater than .5, (\( \varphi_{ij} > .5 \)), the probability of failure is less than .5 (1-\( \varphi_{ij} < .5 \)), the odds of success are greater than 1 and the logit is positive (\( \eta_{ij} > 0 \)). Note that while \( \varphi_{ij} \) is constrained to be in the interval (0, 1), the odds can assume any value from 0 to infinity, and the logit, \( \eta_{ij} \), can take any real value. The *logarithm* transforms a multiplicative to an additive scale and then the set of positive real numbers to the whole real line.
Figure 4: The logit function, \( \text{logit}(\phi_{ij}) = \log \left( \frac{\phi_{ij}}{1 - \phi_{ij}} \right) \), where \( \phi_{ij} \) is the probability of success on each trial.

The transformed predicted value \( \eta_{ij} \) is related to the predictors of the model through the linear structural model:

\[
\eta_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \ldots + \beta_{pj}X_{ pij}
\]  (1.4)

where \( X_{1ij}, \ldots, X_{ pij} \) are the level-1 predictors and \( \beta_{0j}, \ldots, \beta_{pj} \) are regression coefficients from level one. Estimates for \( \beta \)s in equation (1.4) generate a predicted log-odds \( \eta_{ij} \) for any case that can be converted to odds by taking in equation (1.3) the \( \exp(\eta_{ij}) \):

\[
\exp(\eta_{ij}) = \frac{\phi_{ij}}{1 - \phi_{ij}}
\]  (1.5)

A predicted log odds can also be converted to a predicted probability \( \phi_{ij} \) from the equation (1.5) after eliminating the denominator.
\[(1 - \varphi_{ij}) \exp \eta_{ij} = \varphi_{ij} \quad (1.6)\]

and solving the equation for \(\varphi_{ij}\)

\[
\exp \eta_{ij} - \varphi_{ij} \exp \eta_{ij} = \varphi_{ij} 
\quad (1.7)
\]

\[
\exp \eta_{ij} = \varphi_{ij} \exp \eta_{ij} + \varphi_{ij} 
\quad (1.8)
\]

\[
\exp \eta_{ij} = \varphi_{ij} (\exp \eta_{ij} + 1) 
\quad (1.9)
\]

\[
\varphi_{ij} = \frac{\exp \eta_{ij}}{\exp \eta_{ij} + 1} 
\quad (1.10)
\]

After dividing the nominator and denominator by \(\exp \eta_{ij}\), the predicted probability \(\varphi_{ij}\) has the following form:

\[
\varphi_{ij} = \frac{1}{1 + \exp(-\eta_{ij})} 
\quad (1.11)
\]

Note that for every value of \(\eta_{ij}\) applying equation 1.5 will produce a value of \(\varphi_{ij}\) between zero and one.

Thus, in multilevel logistic regression, outcome is modeled by a linear relationship based on log odds. Three pieces of information are important for reporting the results: (1) the outcome expressed in log odds (equations 1.3 and 1.4); (2) the corresponding odds ratio (equation 1.5); and (3) the corresponding predicted probability of an event (equation 1.11).

The level-2 model in GLMM has the following form:

\[
\beta_{ij} = \gamma_{qo} + \sum_{s=1}^{S_i} \gamma_{qi} W_{sj} + u_{qj},
\quad (1.12)
\]
where $W_{sj}$ are the level-2 predictors, $\gamma_{q0}$, \ldots, $\gamma_{qs}$ are the regression coefficients at level 2 and $u_{qj}$ are the random effects with $q=0, \ldots, Q$ having a multivariate normal distribution with component means of zero and variance-covariance matrix $T$ where

$$T = \begin{pmatrix} \tau_{00} & \tau_{10} & \tau_{11} \\ \tau_{10} & \tau_{00} & \tau_{11} \\ \vdots & \vdots & \vdots \\ \tau_{Q0} & \tau_{Q1} & \cdots & \tau_{QQ} \end{pmatrix}$$

(1.13)

and $\tau_{00} = \text{Var}(u_{0j})$, $\tau_{11} = \text{Var}(u_{1j})$, \ldots, $\tau_{QQ} = \text{Var}(u_{Qj})$, $\tau_{10} = \text{Cov}(u_{1j},u_{0j})$, $\tau_{Q0} = \text{Cov}(u_{Qj},u_{0j})$, $\tau_{Q1} = \text{Cov}(u_{Qj},u_{1j})$ are the variances and covariances among the level-2 random effects. The traditional approach to analyzing dichotomous outcomes without a logit link cannot make any assumption regarding the random effects. Since there are only two possible outcomes for the dependent variables, the errors from the model that seek to predict probability of success are not normally distributed. In fact, the errors are heteroscedastic, and the variance itself depends on the predicted values.

I conducted the GLMM in a sequence of steps that lead up to the final model. I screened the numerous potential predictors through linear regression to eliminate those that were not significant to prediction. There are two exploratory strategies available in MLM. Hox (2002) presented an exploratory strategy where the equations and the computer analyses are divided into different models of increasing complexity. The second strategy, considered the more efficient, is a top-down approach (Tabachnick & Fidell, 2007) where the starting model is the most complex one, and includes random effects, higher-level predictors and cross-level interactions. In the next steps the nonsignificant effects are successively eliminated. A drawback of the second method is that, in some instances, the iterative process might fail to converge to an admissible
solution due to the complexity of the model. In my analysis I used Tabachnick and Fidell’s (2007) exploratory strategies.

In general, a two-level model embodies two types of research questions: level-1 questions about *within-person* change and level-2 questions about *between-person* differences in change.

The level-1 component of the multilevel model represents the change which one expects that each member of the population may experience in different classes/sections. I expected that whatever level-1 submodel was specified, the observed data would come from a population where the model was functioning. Thus I inspected visually the empirical growth plots to meet this expectation. When examining empirical growth plots, one must make decisions about the type of the model (linear or curvilinear, smooth or jagged, continuous or disjoint), and about the need for second level modeling. Figures 5, 6, and 7 show a preliminary analysis of the second level regression for each of the four classes (sections) analyzed in the project and the two Level-1 predictors, GENDER, FIRST TRY RATIO and TOO HARD RATIO.
Figure 5. First order regression with the covariate GENDER for the four classes analyzed.

Figure 6. First order regression with the covariate FIRST TRY RATIO for the four classes analyzed.
Figure 7. First order regression with the covariate TOO HARD RATIO for the four classes analyzed.

By visual inspection I determined that:

a) Differences in the slope values between classes were present.

b) Location of the means for the dependent variable (SUCCESS RATIO) varied with the class.

c) The data distribution in each class followed an approximate linear distribution.

Thus, I concluded that there is a need for a second level modeling.

In GLMM a number of decisions have to be made prior to formulation of any specific model. After determining which predictor to include in the model, the researcher must decide to fix or let random the value of each specific parameter. That means that a variable could be considered as a fixed effect over all groups or allowed to vary over the groups (random effect). For relatively small group sizes (ranging from 2 to 100) the data do not contain much information about the main effects of the groups and many parameters have large standard errors (Snijders & Bosker, 2011). In this case the overfitting is avoided by using the random coefficient model. The decision is made
separately for each level-1 predictor. The level-2 predictors may not be considered random because there is no higher level within which they can vary (Tabachnick & Fidell, 2007). The assumption for the random coefficient model is that the random coefficients are normally distributed. If this assumption is a poor approximation, the obtained results may be unreliable (Snijders & Bosker, 2011). For this reason I checked the assumption of normality of the random effects on the first and final model.

**Summary of the Considered Models**

I analyzed and categorized 11 models based on their goodness of fit to determine the best model for the collected data. The first model, called the unconditional means model, contains no predictors at the first or second level except for the group intercept. A significant value for the intraclass correlation parameter determined by this model justifies the use of the multilevel modeling. The next two models (Model 2 and 3) each contained only one indicator at the second Level (ELL SECTION and SCHOOL, respectively). I constructed these models to verify if the indicators ELL SECTION and SCHOOL were relevant for the analysis. The following three models (Model 4, 5, and 6) contained only one Level one indicator each (FIRST TRY RATIO, TOO HARD RATIO, AND GENDER, respectively). As in the case of the previous two models, I constructed these models to determine the need for the inclusion of the parameters FIRST TRY RATIO, TOO HARD RATIO, AND GENDER in the final model. Once all the possible indicators had been identified, I constructed Model 7 by including them together with all possible interactions between terms at level one and level two. I examined this most general model and constructed successively the next four models (Model 8, 9, 10 and 11).
by eliminating the least significant interaction. I established the level of significance for each independent term or interaction based on the probability of a t-value larger than the one corresponding to that specific term. The higher the value of this probability, the less significant the term. I considered p=.05 as an acceptable value for this probability. At the end, I examined the last five models for the goodness of fit.

**Detailed Presentation of the Investigated Models**

**Model 1: Unconditional Means Model**

The first model is an intercepts-only model ("null model"), also known as the logistic-empty model or the unconditional-means model, in which there are no predictors at either child or class level and where the response represented by the dependent variable score for an individual is predicted only by an intercept that varies across groups. The unconditional-means model provides an overall estimate of the likelihood of success for this sample, as well as information about the variability in the probability of success between classes. The following equations define each level of the unconditional means model:

Level 1 (Student level): \( \eta_{ij} = \beta_{0j} \)

Level 2 (Classroom level): \( \beta_{0j} = \gamma_{00} + u_{0j}, \quad u_{0j} \sim N(0, \tau_{00}) \)

Combined: \( \eta_{ij} = \gamma_{00} + u_{0j} \)

where

- \( \eta_{ij} \) is the log odds of success of student \( i \) in class/section \( j \)
- \( \beta_{0j} \) is the log odds of success for students in \( j \)th class/section
- \( \gamma_{00} \) is the average log odds of success across all classes/sections
$u_{0j}$ is the unique effect of class/section $j$ on average log odds of success (intercept) 

$\tau_{00}$ is the variance between classes/sections in class/section-average log odds of success.

The assumption $u_{0j} \sim N(0, \tau_{00})$ for the level 2 is that the $u_{0j}$ random effects are normally distributed with a mean of zero and with variability $\text{var}(u_{0j})=\tau_{00}$.

The syntax for the unconditional means model in SAS 9.3 is:

```
proc glimmix empirical data=temp method=RSPL;
   class Section;
   model Success_Ratio = /s link=logit dist=binomial;
   random intercept /s type = vc sub=Section;
   ODS OUTPUT CovParms=Cov;
run;
```

The PROC GLIMMIX statement invokes the procedure. The EMPIRICAL (sandwich) estimators make the analysis robust against misspecification of the covariance structure and adjust for small-sample bias. The default-estimation method is RSPL, a pseudo-likelihood method used in GLMM. The CLASS statement instructs the GLIMMIX procedure to treat classes/sections (labeled here as SECTION) as classification variables. The MODEL instruction declares SUCCESS RATIO to be the dependent variable (DV). Note that there is nothing after the equal sign because there are no predictors in a logistic-empty model, except for the global intercept ($\gamma_{00}$) which in the SAS software is taken into account automatically. The SOLUTION option in the MODEL statement (“/s”) requests that solutions for the fixed effects (parameter estimates) be displayed. This request provides parameter estimates and significance test
for fixed effects. The GLIMMIX procedure selects the binomial distribution as the
response distribution. Once the distribution is determined, the procedure selects the link
function for the model. The RANDOM instruction sets the group intercept
(classes/sections) to be random. The “type= vc” (variance components) option is the
default structure which invokes the variance components structure, allowing for a
different variance component for each random effect, but sets the covariances to zero.
Research units “sub”(SUBJECT) for the random part of the model are the
classes/sections labeled SECTIONS, which are identified as CLASS. This categorical
variable indicates how level-2 units (SECTIONS) are formed from the level 1 units
(students). At the end, the ODS OUTPUT command instructs SAS to print the output,
including the covariance parameter estimates. The RUN statement commands SAS to
run the identified code.

Model 2: One Level-2 Predictor (ELL SECTION), no Level-1 Predictors

The following equations define each level of the model with one predictor, ELL
SECTION, at the second level:

Level 1 (Student level): \( \eta_{ij} = \beta_{0j} \)
Level 2 (Classroom level): \( \beta_{0j} = \gamma_{00} + \gamma_{01} \text{ ELL SECTION}_{ij} + u_{0j}, \)
\( u_{0j} \sim N (0, \tau_{00}) \)
Combined: \( \eta_{ij} = \gamma_{00} + \gamma_{01} \text{ ELL SECTION}_{ij} + u_{0j} \)

where

\( \eta_{ij}, \beta_{0j}, \gamma_{00}, u_{0j}, \tau_{00} \) are defined like before and
\( \gamma_{01} \) is the overall regression coefficient for the relationship (slope) between log odds of success and ELL SECTIONs.

The syntax for this model in SAS 9.3 is:

```sas
proc glimmix empirical data= temp method=RSPL;
   class Section;
   model Success_Ratio = ELL_Section/s link=logit dist=binomial;
   random intercept/s type = vc sub=Section;
run;
```

The SAS GLIMMIX model instruction declares SUCCESS RATIO to be the outcome with ELL SECTION as a fixed-effect predictor at level 2. This is reflected in the MODEL line, and the remaining syntax is the same as the intercept-only model.

**Model 3: One Level-2 Predictor (SCHOOL), no Level-1 Predictors**

The following equations define each level of the model with one predictor, SCHOOL, at the second level:

- **Level 1 (Student level):** \( \eta_{ij} = \beta_{0j} \)
- **Level 2 (Classroom level):** \( \beta_{0j} = \gamma_{00} + \gamma_{02} \text{SCHOOL}_{ij} + u_{0j} \), \( u_{0j} \sim N(0, \tau_{00}) \)
- **Combined:** \( \eta_{ij} = \gamma_{00} + \gamma_{02} \text{SCHOOL}_{ij} + u_{0j} \)

where

- \( \eta_{ij}, \beta_{0j}, \gamma_{00}, u_{0j}, \tau_{00} \) are defined like before and
- \( \gamma_{02} \) is the overall regression coefficient for the relationship (slope) between log odds of success and SCHOOL.

The syntax for this model in SAS 9.3 is:
The SAS GLIMMIX model instruction declares SUCCESS RATIO to be the outcome with SCHOOL as a fixed-effect predictor at level 2. This is reflected in the MODEL line, and the remaining syntax is the same as the intercept-only model.

**Model 4: One Level-1 Effect (FIRST TRY RATIO), no Level-2 Covariate**

The following equations define each level of the model with one predictor, FIRST TRY RATIO, at the first level:

**Level 1 (Student level):**  
\[ \eta_{ij} = \beta_{0j} + \beta_{1j} \text{FIRST TRY RATIO}_{ij} \]

**Level 2 (Classroom level):**  
\[ \beta_{0j} = \gamma_{00} + u_{0j}, \]
\[ \beta_{1j} = \gamma_{10} + u_{1j}, \]

**Combined:**  
\[ \eta_{ij} = \gamma_{00} + \gamma_{10} \text{FIRST TRY RATIO}_{ij} + [u_{1j} \text{FIRST TRY RATIO}_{ij} + u_{0j}] \]

where

- \( \eta_{ij}, \beta_{0j}, \gamma_{00}, u_{0j} \) are defined like before and
- \( \beta_{1j} \) is the slope for the relationship in \( j \)th class/section between log odds of success and the level-1 predictor (FIRST TRY RATIO)
- \( \gamma_{10} \) is the overall regression coefficient for the relationship (slope) between log odds of success and FIRST TRY RATIO
u_{ij} is the unique effect of class/section j on slope when the value of level-2 predictors are zero.

The syntax for this model in SAS 9.3 is:

```sas
proc glimmix empirical data= temp method=RSPL;
  class Section;
  model Success_Ratio = First_Try_Ratio/solution link=logit dist=binomial;
  random intercept First_Try_Ratio/s type = vc sub=Section;
run;
```

The SAS GLIMMIX model instruction declares SUCCESS RATIO to be the outcome with FIRST TRY RATIO as a fixed and random predictor at level 1. This is reflected in the MODEL and RANDOM line, and the remaining syntax is the same as the intercept-only model.

**Model 5: One Level-1 Effect (TOO HARD RATIO), no Level-2 Covariate**

The following equations define each level of the model with one predictor, TOO HARD RATIO, at the first level:

**Level 1 (Student level):** \( \eta_{ij} = \beta_{0j} + \beta_{2j} \text{ TOO HARD RATIO}_{ij} \)

**Level 2 (Classroom level):** \( \beta_{0j} = \gamma_{00} + u_{0j} \), \( \beta_{2j} = \gamma_{20} \),

**Combined:** \( \eta_{ij} = \gamma_{00} + \gamma_{20} \text{ TOO HARD RATIO}_{ij} + u_{0j} \)

where

\( \eta_{ij}, \beta_{0j}, \gamma_{00}, u_{0j} \) are defined like before and
\( \beta_{2j} \) is the slope for the relationship in \( j \)th class/section between log odds of success and the level-1 predictor (TOO HARD RATIO).

\( \gamma_{20} \) is the overall regression coefficient for the relationship (slope) between log odds of success and TOO HARD RATIO.

The syntax for this model in SAS 9.3 is:

```
proc glimmix empirical data= temp method=RSPL;
  class Section;
  model Success_Ratio = Too_Hard_Ratio/solution link=logit dist=binomial;
  random intercept /s type = vc sub=Section;
run;
```

The SAS GLIMMIX model instruction declares SUCCESS RATIO to be the outcome with TOO HARD RATIO as a fixed predictor at level 1. This is reflected in the MODEL line, and the remaining syntax is the same as the intercept-only model.

**Model 6: One Level-1 Effect (GENDER), no Level-2 Covariate**

The following equations define each level of the model with one predictor, GENDER, at the first level:

Level 1 (Student level):

\[
\eta_{ij} = \beta_0 + \beta_3 \text{GENDER}_{ij}
\]

Level 2 (Classroom level):

\[
\beta_0 = \gamma_{00} + u_{0j}, \quad \beta_3 = \gamma_{30},
\]

Combined:

\[
\eta_{ij} = \gamma_{00} + \gamma_{30} \text{GENDER}_{ij} + u_{0j}
\]

where

\( \eta_{ij}, \beta_{0j}, \gamma_{00}, u_{0j} \) are defined like before and
\( \beta_{3j} \) is the slope for the relationship in \( j \)th class/section between log odds of success and the level-1 predictor (GENDER)

\( \gamma_{30} \) is the overall regression coefficient for the relationship (slope) between log odds of success and GENDER

The syntax for this model in SAS 9.3 is:

```sas
proc glimmix empirical data= temp method=RSPL;
    class Section;
    model Success_Ratio = Gender/solution link=logit dist=binomial;
    random intercept/s type = vc sub=Section;
run;
```

The SAS GLIMMIX model instruction declares SUCCESS RATIO to be the outcome with GENDER as a fixed predictor at level 1. This is reflected in the MODEL line, and the remaining syntax is the same as the intercept-only model.

**Model 7: General Model with Level-1 Predictors (FIRST TRY RATIO, TOO HARD RATIO, and GENDER), and level-2 Predictors (ELL SECTION and SCHOOL)**

The following equations define each level of the model with three level-1 predictors, FIRST TRY RATIO, TOO HARD RATIO, GENDER, and two level-2 predictors, ELL SECTION and SCHOOL:

Level 1 (Student level):
\[
\eta_{ij} = \beta_{0j} + \beta_{1j} \text{FIRST TRY RATIO}_{ij} + \beta_{2j} \text{TOO HARD RATIO}_{ij} + \beta_{3j} \text{GENDER}_{ij}
\]

Level 2 (Classroom level):
\[
\beta_{0j} = \gamma_{00} + \gamma_{01} \text{ELL SECTION}_j + \gamma_{02} \text{SCHOOL}_j + u_{0j},
\]
\[ \beta_{1j} = \gamma_{10} + \gamma_{11} \text{ELL SECTION}_j + \gamma_{12} \text{SCHOOL}_j + u_{1j}, \]

\[ \beta_{2j} = \gamma_{20} + \gamma_{21} \text{ELL SECTION}_j + \gamma_{22} \text{SCHOOL}_j + u_{2j}, \]

\[ \beta_{3j} = \gamma_{30} + \gamma_{31} \text{ELL SECTION}_j + \gamma_{32} \text{SCHOOL}_j + u_{3j}, \]

Combined:

\[ \eta_{ij} = \gamma_{00} + \gamma_{01} \text{ELL SECTION}_j + \gamma_{10} \text{FIRT TRY RATIO}_i + \gamma_{02} \text{SCHOOL}_j + \gamma_{20} \text{TOO HARD RATIO}_i + \gamma_{30} \text{GENDER}_ij + \gamma_{11} \text{ELL SECTION}_j \times \text{FIRT TRY RATIO}_i + \gamma_{12} \text{SCHOOL}_j \times \text{FIRT TRY RATIO}_i + \gamma_{21} \text{ELL SECTION}_j \times \text{TOO HARD RATIO}_i + \gamma_{22} \text{SCHOOL}_j \times \text{TOO HARD RATIO}_i + \gamma_{31} \text{ELL SECTION}_j \times \text{GENDER}_ij + \gamma_{32} \text{SCHOOL}_j \times \text{GENDER}_ij + [u_{0j} + u_{1j} \text{FIRT TRY RATIO}_i + u_{2j} \text{TOO HARD RATIO}_i + u_{3j} \text{GENDER}_ij] \]

where

\[ \eta_{ij}, \beta_{0j}, \beta_{1j}, \beta_{2j}, \beta_{3j}, \gamma_{00}, \gamma_{10}, \gamma_{20}, \gamma_{30}, \gamma_{01}, \gamma_{02}, \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \gamma_{31}, \gamma_{32} \] are defined like before and

\[ \gamma_{11} \] is the mean difference in FIRT TRY RATIO- log odds of success slopes among ELL SECTIONs

\[ \gamma_{12} \] is the mean difference in FIRT TRY RATIO- log odds of success slopes between SCHOOLs

\[ \gamma_{21} \] is the mean difference in TOO HARD RATIO - log odds of success slopes among ELL SECTIONs

\[ \gamma_{22} \] is the mean difference in TOO HARD RATIO - log odds of success slopes between SCHOOLs

\[ \gamma_{31} \] is the mean difference in GENDER - log odds of success slopes among ELL SECTIONs
\( \gamma_{32} \) is the mean difference in GENDER - log odds of success slopes between SCHOOLs

\( u_{2j} \) is the unique effect of group \( j \) on slope when the value of level-2 predictors are zero

\( u_{3j} \) is the unique effect of group \( j \) on slope when the value of level-2 predictors are zero

The syntax for this model in SAS 9.3 is:

```
proc glimmix empirical data= temp method=RSPL;
  class Section;
  model Success_Ratio = First_Try_Ratio Too_Hard_Ratio Gender ELL_Section School
          First_Try_Ratio*ELL_Section First_Try_Ratio*School
          Too_Hard_Ratio*ELL_Section Too_Hard_Ratio*School
          Gender*ELL_Section Gender*School/s link=logit dist=binomial;
  random intercept First_Try_Ratio /s type = vc sub=Section;
run;
```

The SAS GLIMMIX model instruction declares SUCCESS RATIO to be the outcome with FIRST TRY RATIO, TOO HARD RATIO, and GENDER as fixed level-1 predictors and ELL SECTION and SCHOOL as fixed level-2 predictors. This is reflected in the MODEL line where the six cross-level interactions between the level-2 predictors and level-1 predictors are also present. In addition, the FIRST TRY RATIO is considered random which is reflected in the RANDOM line. The remaining syntax is the same as the intercept-only model.

For many observed count data, as in this study, it is common to have the sample variance smaller than the sample mean which are referred to as underdispersion. The cause of this phenomenon is clustering in the population. McCullagh and Nelder (1989)
point out that the dispersion parameter depends on the cluster size and on the variability of the probability of success from cluster to cluster but not on the sample size. To remedy these situations, the authors suggest that the model should include a variance function that does not correspond to any of the built-in distributions. In GLIMMIX, the analysis of the revised model is obtained with the statement “_variance_ = _mu_ **2 * (1-_mu_)**2” (SAS Institute, 2006). Such dispersion components do not affect the parameter estimates, only their standard errors. A genuine random effect, on the other hand, affects both the parameter estimates and their standard errors.

Model 8: The General Model with One Interaction Removed (TOO HARD RATIO and SCHOOL)

The following equations define each level of the model with three level-1 predictors, FIRST TRY RATIO, TOO HARD RATIO, GENDER, and two level-2 predictors, ELL SECTION and SCHOOL with the intersection between TOO HARD RATIO and SCHOOL removed from the General Model (Model 7):

Level 1(Student level): \[ \eta_{ij} = \beta_0 + \beta_1 \text{FIRST TRY RATIO}_{ij} + \beta_2 \text{TOO HARD RATIO}_{ij} + \beta_3 \text{GENDER}_{ij} \]

Level 2 (Classroom level): \[ \beta_0 = \gamma_{00} + \gamma_{01} \text{ELL SECTION}_j + \gamma_{02} \text{SCHOOL}_j + u_{0j}, \]
\[ \beta_1 = \gamma_{10} + \gamma_{11} \text{ELL SECTION}_j + \gamma_{12} \text{SCHOOL}_j + u_{1j}, \]
\[ \beta_2 = \gamma_{20} + \gamma_{21} \text{ELL SECTION}_j + u_{2j}, \]
\[ \beta_3 = \gamma_{30} + \gamma_{31} \text{ELL SECTION}_j + \gamma_{32} \text{SCHOOL}_j + u_{3j}, \]

Combined: \[ \eta_{ij} = \gamma_{00} + \gamma_{01} \text{ELL SECTION}_j + \gamma_{10} \text{FIRST TRY RATIO}_{ij} + \gamma_{12} \text{SCHOOL}_j + \gamma_{20} \text{TOO HARD RATIO}_{ij} + \gamma_{30} \text{GENDER}_{ij} + \gamma_{11} \text{ELL SECTION}_j \times \text{FIRST TRY RATIO}_{ij} + \gamma_{12} \]
\[
\text{SCHOOL}_j \times \text{FIRST TRY RATIO}_{ij} + \gamma_{21} \times \text{ELL SECTION}_j \times \text{TOO HARD RATIO}_{ij} + \gamma_{31} \times \text{ELL SECTION}_j \times \text{GENDER}_{ij} \\
+ \gamma_{32} \times \text{SCHOOL}_j \times \text{GENDER}_{ij} + [u_{0j} + u_{1j} \times \text{FIRST TRY RATIO}_{ij} + u_{2j} \times \text{TOO HARD RATIO}_{ij} + u_{3j} \times \text{GENDER}_{ij}]
\]

where all variables are defined like before.

The syntax for this model in SAS 9.3 is:

```sas
proc glimmix empirical data= temp method=RSPL;
   class Section;
   _variance_ = ( _mu_ * (1-_mu_) )**2;
   model Success_Ratio = First_Try_Ratio Too_Hard_Ratio Gender ELL_Section School
                           First_Try_Ratio*ELL_Section First_Try_Ratio*School
                           Too_Hard_Ratio*ELL_Section
                           Gender*ELL_Section Gender*School/s link=logit dist=binomial;
   random intercept/s type = vc sub=Section;
run;
```

I eliminated in this model the interaction between TOO HARD RATIO and SCHOOL, as reflected in the MODEL line from the SAS syntax. The attempt of dropping the second interaction between TOO HARD RATIO and ELL SECTION lead to the result that TOO HARD RATIO is not significant, contradicting the results from the model 5. I also removed the random effect of the variable FIRST TRY RATIO for a better fit of the data.
Model 9: The General Model with Two Interactions Removed (FIRST TRY RATIO - ELL SECTION and TOO HARD RATIO - SCHOOL)

The following equations define each level of the model with three level-1 predictors, FIRST TRY RATIO, TOO HARD RATIO, GENDER, and two level-2 predictors, ELL SECTION and SCHOOL with the intersection between FIRST TRY RATIO and ELL SECTION removed from the previous model:

Level 1 (Student level):
\[ \eta_{ij} = \beta_{0j} + \beta_{1j} \text{FIRST TRY RATIO}_{ij} + \beta_{2j} \text{TOO HARD RATIO}_{ij} + \beta_{3j} \text{GENDER}_{ij} \]

Level 2 (Classroom level):
\[ \beta_{0j} = \gamma_{00} + \gamma_{01} \text{ELL SECTION}_j + \gamma_{02} \text{SCHOOL}_j + u_{0j}, \]
\[ \beta_{1j} = \gamma_{10} + \gamma_{12} \text{SCHOOL}_j + u_{1j}, \]
\[ \beta_{2j} = \gamma_{20} + \gamma_{21} \text{ELL SECTION}_j + u_{2j}, \]
\[ \beta_{3j} = \gamma_{30} + \gamma_{31} \text{ELL SECTION}_j + \gamma_{32} \text{SCHOOL}_j + u_{3j}, \]

Combined:
\[ \eta_{ij} = \gamma_{00} + \gamma_{01} \text{ELL SECTION}_j + \gamma_{10} \text{FIRST TRY RATIO}_{ij} + \gamma_{02} \text{SCHOOL}_j + \gamma_{12} \text{SCHOOL}_j \times \text{FIRST TRY RATIO}_{ij} + \gamma_{20} \text{TOO HARD RATIO}_{ij} + \gamma_{30} \text{GENDER}_{ij} + \gamma_{31} \text{ELL SECTION}_j \times \text{TOO HARD RATIO}_{ij} + \gamma_{32} \text{SCHOOL}_j \times \text{GENDER}_{ij} + u_{0j} + u_{1j} \text{FIRST TRY RATIO}_{ij} + u_{2j} \text{TOO HARD RATIO}_{ij} + u_{3j} \text{GENDER}_{ij} \]

where all variables are defined like before.

The syntax for this model in SAS 9.3 is:

```
proc glimmix empirical data= temp method=RSPL;
   class Section;
   _variance_ = ( _mu_ * (1-_mu_))**2;
```
model Success_Ratio = First_Try_Ratio Too_Hard_Ratio Gender ELL_Section School
First_Try_Ratio*School Too_Hard_Ratio*ELL_Section
Gender*ELL_Section Gender*School/s link=logit dist=binomial;
random intercept/s type = vc sub=Section;
run;

I eliminated in this model the interaction between FIRST TRY RATIO and ELL SECTION, as reflected in the MODEL line from the SAS syntax.

Model 10: The General Model with Three Interactions Removed (FIRST TRY RATIO - ELL SECTION, TOO HARD RATIO - SCHOOL, and TOO HARD RATIO - ELL SECTION)

The following equations define each level of the model with three level-1 predictors, FIRST TRY RATIO, TOO HARD RATIO, GENDER, and two level-2 predictors, ELL SECTION and SCHOOL with the intersection between TOO HARD RATIO and ELL SECTION removed from the previous model:

Level 1 (Student level): \( \eta_{ij} = \beta_{0j} + \beta_{1j} \text{FIRST TRY RATIO}_{ij} + \beta_{2j} \text{TOO HARD RATIO}_{ij} + \beta_{3j} \text{GENDER}_{ij} \)

Level 2 (Classroom level): \( \beta_{0j} = \gamma_{00} + \gamma_{01} \text{ELL SECTION}_{j} + \gamma_{02} \text{SCHOOL}_{j} + u_{0j}, \)
\( \beta_{1j} = \gamma_{10} + \gamma_{12} \text{SCHOOL}_{j} + u_{1j}, \)
\( \beta_{2j} = \gamma_{20} + u_{2j}, \)
\( \beta_{3j} = \gamma_{30} + \gamma_{31} \text{ELL SECTION}_{j} + \gamma_{32} \text{SCHOOL}_{j} + u_{3j}, \)

Combined: \( \eta_{ij} = \gamma_{00} + \gamma_{01} \text{ELL SECTION}_{j} + \gamma_{10} \text{FIRST TRY RATIO}_{ij} + \)
\[ \gamma_{02} \text{SCHOOL}_{ij} + \gamma_{20} \text{TOO HARD RATIO}_{ij} + \gamma_{30} \text{GENDER}_{ij} \]

\[ + \gamma_{12} \text{SCHOOL}_{ij} \times \text{FIRT TRY RATIO}_{ij} + \gamma_{31} \text{ELL SECTION}_{ij} \times \text{GENDER}_{ij} + \]

\[ u_{0j} + u_{ij} \text{FIRT TRY RATIO}_{ij} + u_{2j} \text{TOO HARD RATIO}_{ij} + \]

\[ u_{3j} \text{GENDER}_{ij} \]

where all variables are defined like before.

The syntax for this model in SAS 9.3 is:

```sas
proc glimmix empirical data= temp method=RSPL;
   class Section;
   _variance_ = ( _mu_ * (1-_mu_) )**2;
   model Success_Ratio = First_Try_Ratio Too_Hard_Ratio Gender ELL_Section School
      First_Try_Ratio*School
      Gender*ELL_Section Gender*School/s link=logit dist=binomial;
   random intercept/s type = vc sub=Section;
run;
```

I eliminated in this model the interaction between TOO HARD RATIO and ELL SECTION, as reflected in the MODEL line from the SAS syntax.

The final model allows for cross-level interaction between FIRST TRY RATIO and SCHOOL and between GENDER and ELL SECTION on success. The variables SCHOOL and ELL SECTION are included as predictors of the intercepts across classes/sections, as well as on the effect of a student’s GENDER and FIRST TRY RATIO characteristics on student-level success outcomes across classes/sections. Thus, the intercepts and the slopes from the level-1 equation are treated as outcomes in a
regression model at level two. The fixed effects in the equation predicting the slopes are referred to as cross-level interactions. They represent the effect of a level-2 variable on a level-1 predictor of success. These effects tell how the relationship between level-1 predictors and the log odds tends to vary across groups, based on the level-2 variables.

**Model 11: The General Model with all Interactions Removed**

The following equations define each level of the model with three level-1 predictors, FIRST TRY RATIO, TOO HARD RATIO, GENDER, and two level-2 predictors, ELL SECTION and SCHOOL and no interactions:

**Level 1 (Student level):**

\[ \eta_{ij} = \beta_0j + \beta_{1j} \text{FIRST TRY RATIO}_{ij} + \beta_{2j} \text{TOO HARD RATIO}_{ij} + \beta_{3j} \text{GENDER}_{ij} \]

**Level 2 (Classroom level):**

\[ \beta_{0j} = \gamma_{00} + \gamma_{01} \text{ELL SECTION}_j + \gamma_{02} \text{SCHOOL}_j + u_{0j}, \]
\[ \beta_{1j} = \gamma_{10} + u_{1j}, \]
\[ \beta_{2j} = \gamma_{20} + u_{2j}, \]
\[ \beta_{3j} = \gamma_{30} + u_{3j}, \]

**Combined:**

\[ \eta_{ij} = \gamma_{00} + \gamma_{01} \text{ELL SECTION}_j + \gamma_{10} \text{FIRST TRY RATIO}_{ij} + \gamma_{20} \text{SCHOOL}_j + \gamma_{30} \text{TOO HARD RATIO}_{ij} + \gamma_{30} \text{GENDER}_{ij} + [u_{0j} + u_{1j} \text{FIRST TRY RATIO}_{ij} + u_{2j} \text{TOO HARD RATIO}_{ij} + u_{3j} \text{GENDER}_{ij}] \]

where all variables are defined like before.

The syntax for this model in SAS 9.3 is:

```sas
proc glimmix empirical data= temp method=RSPL;
   class Section;
```
\_variance_ = ( \_mu_ \* (1 - \_mu\_) \^2; \\

_model Success\_Ratio = First\_Try\_Ratio Too\_Hard\_Ratio Gender ELL\_Section School /s link=\text{logit} dist=\text{binomial}; \\

_random intercept/s type = vc sub=Section; \\

run;

I removed in this model all interactions, as reflected in the MODEL line from the SAS syntax.

Previously, I included the variables SCHOOL and ELL SECTION as predictors of the intercepts across classes/sections. Thus, the intercepts and the slopes from the level-1 equation are treated as outcomes in a regression model at level two. The fixed effects in the equation predicting the slopes are referred to as cross-level interactions. They represented the effect of a level-2 variable on a level-1 predictor of success. These effects tell how the relationship between level-1 predictors and the log odds tends to vary across groups, based on the level-2 variables.

\textbf{Model Comparison}

The Fit Statistics table from the output provides statistics about each estimated model. In the presence of random effects and a conditional binomial distribution, PROC GLIMMIX does not use maximum likelihood for estimation. For a model containing random effects, the GLIMMIX procedure, by default, estimates the parameters by applying pseudo-likelihood techniques. According to the SAS Institute (2006), there are two possible groups of estimators. If the estimation method permits the true log likelihood or residual log likelihood, the description of the first entry reads accordingly,
and it corresponds to the negative of twice the (possibly restricted) log likelihood, log pseudo-likelihood, or log quasi-likelihood. Otherwise, the fit statistics are preceded by the words Pseudo- or Quasi-, for Pseudo- and Quasi-Likelihood estimation, respectively.

In the first case, in general, a value of -2 times the log likelihood (-2LL) for a particular fitted model can assess the quality of fit for comparison of competing models. The value -2LL is also called the deviance of the model and represents how poorly the model fits the data. If deviance is reduced by a competing nested model, the competing model is preferred. When the models are nested (i.e., one model is the extension of another model) the difference in deviance follows a chi-square distribution with degree of freedom determined by the difference in number of estimated parameters, and it can be compared statistically.

In the second case, according to the SAS Institute (2006), the (residual) log pseudo-likelihood is reported. The (residual) log pseudo-likelihood in a GLMM is the (residual) log likelihood of a linearized model for an approximated model, and it is not possible to compare these values across different statistical models, even if the models are nested. The generalized chi-square statistic is a quadratic form in the marginal residuals that takes correlations among the data into account and measures the residual sum of squares in the final model. The ratio with its degrees of freedom is a measure of variability of the observation about the mean model, and this is the only fit statistics that I will focus on. A value close to 1 indicates that the variability in these data has been properly modeled, and that there is no residual underdispersion/overdispersion.
CHAPTER 4

FINDINGS

In this chapter, I present the results of the analysis that I conducted with two-level GLMM for the final model that represented the best fit for the actual data set. First, I present the descriptive results. It is important to examine these statistics to determine whether enough variation exists in dependent measures. Next, I provide the demonstration of the best model. Finally, I present the findings of the final model that represents the best fit for the data I analyzed.

Descriptive Results

I present descriptive statistics for the data aggregated across grade level in Table 2. The outcome variable of interest is the SUCCESS RATIO. The typical 7th grader, on average, had a rate of success of 72% of the total number of problems. The dummy variable from level-1, gender, indicated that 46% of the sample was male (MALE=1). The other two level-1 predictors, FIRST TRY RATIO (the ratio of problems solved correctly from the first attempt over the total number of problems solved correctly) and TOO HARD RATIO (the ratio of problems considered by each student to be too hard over the total number of problems presented to the student) captured the student behavior when solving the administered word problems. The mean FIRST TRY RATIO across 84 students (45%), and the mean TOO HARD RATIO (10%) both with small standard deviation showed student behavioral consistency in solving from the first attempt and rating problems.
The variables TOO HARD RATIO and FIRST TRY RATIO were centered on individual grand mean values. Centering is a process of changing a raw score to a deviation score by subtracting a mean value from each predictor score so that each variable has a mean of zero. The reason for doing this is to prevent multicollinearity when predictors are components of interactions and to simplify interpretations. Centering is most commonly performed on level-1 predictors when there is no meaning to a value of zero on a predictor. Level-2 predictors are not usually centered unless interactions are formed among two or more continuous level-2 predictors, and centering outcomes is unusual because of the difficulty of the interpretations. In addition, if interactions are formed, predictors in the interactions are bound to be correlated with the main effects. In this study, I used centering to solve the problem of multicollinearity between interactions and their main effects (Tabachnick & Fidell, 2007).

I found two predictors of interest at level two: ELL SECTION, an indicator of the ratio of ELs in each of the four classes/sections, and a dummy variable, SCHOOL, capturing the distribution of the students in schools. I sampled students from two schools (LCMS=1, AMS=0), and four classes (three classes from LCMS and one class from AMS). I coded the class/section with the highest ELs ratio, which represents the AMS, with 0. I coded the rest of three classes/sections from the LCMS with 1, 2 and 3 in decreasing order of the ratio of ELs in each of the classes/sections. Seventy-one percent of students were from LCMS. The number of ELs ranged from 6% to 42% per class. There was less variation in variables scores at level-1 and level-2 reflected by standard deviations less than 1.00.
Table 2.

Descriptive Statistics for Variables Included in GLMM Analyses

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Try Ratio</td>
<td>84</td>
<td>.45</td>
<td>.15</td>
<td>.11</td>
<td>.80</td>
</tr>
<tr>
<td>Too Hard Ratio</td>
<td>84</td>
<td>.10</td>
<td>.11</td>
<td>0</td>
<td>.47</td>
</tr>
<tr>
<td>Gender</td>
<td>84</td>
<td>.46</td>
<td>.50</td>
<td>0</td>
<td>1.00</td>
</tr>
<tr>
<td>Success Ratio</td>
<td>84</td>
<td>.72</td>
<td>.16</td>
<td>.25</td>
<td>.98</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>J</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELL Section</td>
<td>4</td>
<td>.31</td>
<td>.13</td>
<td>.06</td>
<td>.42</td>
</tr>
<tr>
<td>School</td>
<td>2</td>
<td>.71</td>
<td>.45</td>
<td>0</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Model 1: Unconditional Means Model

Results of the unconditional-means model in terms of estimated logits and corresponding odds ratio are in Table 3.
In this model the predicted logit for a typical class/section was given by the intercept, $\gamma_{00}= 1.07$ (se= .16), also known as the grand mean. This average logit was statistically different from zero, $t(3)=6.64$, $p<.05$. The intercept $\beta_{0j}$ for every class/section was not likely to have a single value as its value depends on the group. This was confirmed by the parameter estimate ($\tau_{00}$) and its standard error provided in the Covariance Parameter Estimates section of the output. The variance in intercepts across classes/sections was $\tau_{00}=.11$ (se= .11), $\chi^2=79.89$, $p<.05$. This showed that the intercept variability with class/section was significant, suggesting the need to consider differences between groups when analyzing the rate of success. Thus, for a class/section with a “typical” rate of success, that is, for a class/section with a random effect $u_{0j} = 0$, the expected log odds of success was 1.07, corresponding to an odds of $\exp \{1.07\}= 2.91$. This corresponded to a probability of $1/ (1+ \exp\{-1.07\}) = .74$.

Under the model’s assumption that the errors at level two followed a normal distribution, I expected the classrooms’ log odds of success, $\beta_{0j}$, to be normally
distributed with a mean of 1.07 and variance $\tau_{00}= .11$. This is the same as saying that the 95% confidence interval of the parameter $\beta_{0j}$ for these classes/sections was between $1.07\pm 1.96 \sqrt{.11} = (.42, 1.72)$. When I converted these log odds to probabilities, 95% of the classrooms were between .60 and .85 with respect to the probability of success. Thus, some classes/sections had a considerable percentage of students who were successful in solving mathematics word problems.

The unconditional-means model is of singular importance in GLMM because it provides information about interclass correlation (ICC), a value helpful in determining whether a multilevel model is required. According to Snijders and Bosker (1999) the formula for interclass correlation is

$$
 ICC = \frac{\tau_{00}}{\tau_{00} + \pi^2 / 3}
$$

where $\pi = 3.14$ and $\tau_{00}$ is the variance between classes/sections in class/section-average log odds of success. The interclass correlation may be interpreted in two equivalent ways: either as the correlation between two randomly drawn individuals in one randomly drawn group, or as the fraction of the total variability that was due to group level where the logistic distribution for the level-1 residual implies a variance of $\pi^2/3 = 3.29$.

The interclass correlation for my sample was

$$
 ICC = \frac{.11}{.11 + 3.29} = .03
$$

indicating that 3% of the total variation in rate of success is attributable to differences among classes/sections. The estimation of the variability attributed to the classes/sections was on the low side compared to other results in educational research that are between
.10 and .25, according to Hedges and Hedberg (2007). However, the ICC value was larger than zero, indicating that a multilevel analysis was appropriate in this case due to a clustering effect. A zero value for ICC implies that there is not a meaningful average difference among classes/sections, and data may be analyzed at the individual (first) level.

**Model 2: One Level-2 Predictor (ELL SECTION), no Level-1 Predictors**

This model considered the effect of EL distribution in each class (i.e., the percentage of ELs in each of the four classes and the manner in which the results at the class level are changing). The Solutions for Fixed Effects section determined that the ELL SECTION slope value was $\gamma_{01} = .73$, statistically significant at the .05 level, and concluded that the percentage of the ELs in each class had a direct influence on the overall results of that class. Based on this, I retained the ELL SECTION in this model. Moreover, the Covariance Parameter Estimates section showed that, indeed, the group intercepts vary across classes/sections.
Table 4

Model 2: One Level-2 Predictor (ELL SECTION), no Level-1 Predictors

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.85</td>
<td>.18</td>
<td>2</td>
<td>4.80</td>
<td>.0408</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELL_Section</td>
<td>.74</td>
<td>.89</td>
<td>80</td>
<td>.83</td>
<td>.0487</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Section</td>
<td>.16</td>
<td>.19</td>
</tr>
</tbody>
</table>

Model 3: One Level-2 Predictor (SCHOOL), no Level-1 Predictors

The Solutions for Fixed Effects section determined that the SCHOOL slope value of $\gamma_2 = .47$ was statistically significant at the .05 level, and concluded that SCHOOL variable should be retained in the model. The Covariance Parameter Estimates section showed that the group intercepts vary across classes/sections.
Table 5

Model 3: One Level-2 Predictor (SCHOOL), no Level-1 Predictors

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.72</td>
<td>.</td>
<td>2</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td>.48</td>
<td>.16</td>
<td>80</td>
<td>2.90</td>
<td>.0048</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Section</td>
<td>.10</td>
<td>.12</td>
</tr>
</tbody>
</table>

Model 4: One Level-1 Effect (FIRT TRY RATIO), no Level-2 Covariate

In this model both the slope and the intercept were allowed to vary with class/section. In order to achieve convergence in Maximum Likelihood, I forced the covariance between slope and intercept to zero. The values of Intercept and Level-1 covariate FIRST TRY RATIO were both significantly different from zero at the .05 level. Thus, I retained the FIRST TRY RATIO variable in this model. The Covariance Parameter Estimates section showed that the FIRST TRY RATIO variable varied across classes/sections.
Table 6

*Model 4: One Level-1 Effect (FIRST TRY RATIO), no Level-2 Covariate*

| Effect             | Estimate | Standard Error | DF  | t Value | Pr > |t| |
|--------------------|----------|----------------|-----|---------|-------|
| Intercept          | 1.15     | .10            | 3   | 11.80   | .0013 |
| First_Try_Ratio    | 3.66     | 1.02           | 3   | 3.58    | .0373 |

* Covariance Parameter Estimates *

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Section</td>
<td>.03</td>
<td>.04</td>
</tr>
<tr>
<td>First_Try_Ratio</td>
<td>Section</td>
<td>4.58</td>
<td>4.27</td>
</tr>
</tbody>
</table>

*Model 5: One Level-1 Effect (TOO HARD RATIO), no Level-2 Covariate*

In this model, I allowed both the slope and the intercept to vary with class/section. The Covariance Parameter Estimates table shows that the standard errors of the intercept and that of TOO HARD RATIO were not significantly different from zero. The values of Intercept and Level-1 covariate TOO HARD RATIO were both significantly different than zero at the .05 level. Thus, I retained the TOO HARD RATIO variable.
Table 7

*Model 5: One Level-1 Effect (TOO HARD RATIO), no Level-2 Covariate*

| Effect               | Estimate | Standard Error | DF | t Value | Pr > |t| |
|----------------------|----------|----------------|----|---------|-------|---|
| Intercept            | 1.07     | .12            | 3  | 9.31    | .0026 |
| Too_Hard_Ratio       | -3.93    | .51            | 3  | -7.67   | .0046 |

**Covariance Parameter Estimates**

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Section</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td>Too_Hard_Ratio</td>
<td>Section</td>
<td>4.30E-18</td>
<td></td>
</tr>
</tbody>
</table>

*Model 6: One Level-1 Effect (GENDER), no Level-2 Covariate*

Here, I allowed both the slope and the intercept to vary with class/section. The Covariance Parameter Estimates table shows that the standard error of the intercept and that of GENDER were not significantly different from zero. The values of Intercept and Level-1 covariate GENDER were both significantly different from zero at the .05 level of significance. Thus, I retained the GENDER variable.
### Table 8

**Model 6: One Level-1 Effect (GENDER), no Level-2 Covariate**

| Effect  | Estimate | Standard Error | DF  | t Value | Pr > |t| |
|---------|----------|----------------|-----|---------|------|---|
| Intercept | 1.19 | .14 | 3 | 8.40 | .0035 |
| Gender | -.25 | .08 | 3 | -3.17 | .0504 |

<table>
<thead>
<tr>
<th>Covariance Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cov Parm</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>Gender</td>
</tr>
</tbody>
</table>

**Model 7: General Model with Level-1 Predictors (FIRST TRY RATIO, TOO HARD RATIO, and GENDER), and level-2 predictors (ELL SECTION and SCHOOL)**

In this model I attempted to include all previously determined significant covariates and their cross-level interactions. The Solution for Fixed Effects section shows that the slope for FIRST TRY RATIO (p=.05), TOO HARD RATIO (p=.05), and GENDER (p=.03) were significantly different from zero at the .05 level. The slope for SCHOOL (p<.003), the interactions between FIRST TRY RATIO and SCHOOL (p<.0001), GENDER and ELL SECTION (p<.0001), and GENDER and SCHOOL (p=.002) were highly significant. The interaction between TOO HARD RATIO and ELL SECTION (p=.90), as well as the interaction between TOO HARD RATIO and SCHOOL (p=.82), were non-significant and eliminated in future analyses. The Covariance Parameter Estimates section shows that the group intercepts and the FIRST
TRY RATIO estimate were zero. This means that there was no unexplained between-
group variability left that could be explained by group-level variable and intercept. The
convergence criterion was met after 17 iterations.
Table 9

*Model 7: General Model with Level-1 Predictors (FIRST TRY RATIO, TOO HARD RATIO, and GENDER), and Level-2 Predictors (ELL SECTION and SCHOOL)*

| Effect                  | Estimate | Standard Error | DF | t Value | Pr > |t|   |
|-------------------------|----------|----------------|----|---------|-------|-----|
| Intercept               | .75      | .08            | 1  | 9.10    | .0697 |
| First_Try_Ratio        | 5.88     | .47            | 1  | 12.47   | .0509 |
| Too_Hard_Ratio         | -2.59    | 1.28           | 70 | -2.03   | .0462 |
| Gender                 | -.11     | .05            | 70 | -2.18   | .0328 |
| ELL_Section            | .60      | .20            | 70 | 3.07    | .0031 |
| School                 | .23      | .07            | 70 | 3.07    | .0031 |
| First_Try*ELL_Section  | .84      | 1.12           | 70 | .74     | .4594 |
| First_Try_Rat*School   | -4.16    | .39            | 70 | -10.75  | <.0001|
| Too_Hard_*ELL_Section  | -.25     | 3.04           | 70 | -.08    | .9345 |
| Too_Hard_Rati*School   | -.25     | 1.11           | 70 | -.23    | .8189 |
| Gender*ELL_Section     | 1.05     | .11            | 70 | 9.11    | <.0001|
| Gender*School          | -.11     | .03            | 70 | -3.23   | .0019 |

<table>
<thead>
<tr>
<th>Covariance Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cov Parm</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>First_Try_Ratio</td>
</tr>
</tbody>
</table>
Model 8: The General Model with One Interaction Removed (TOO HARD RATIO and SCHOOL)

In this model, the Solution for Fixed Effects section shows that the Intercept (p=.03), the slope for FIRST TRY RATIO (<.0001), the slope for TOO HARD RATIO (<.0001), the slope for SCHOOL (<.0001), the interactions between FIRST TRY RATIO and SCHOOL (<.0001), and GENDER and ELL SECTION (<.0001), were significantly different from zero at .05 level of significance. The interaction between FIRST TRY RATIO and ELL SECTION (.47), as well as the interaction between TOO HARD RATIO and ELL SECTION (.64), and GENDER and SCHOOL (.06) were non-significant, and I eliminated them in future analyses. The Covariance Parameter Estimates shows that there was no unexplained variability in this model.
Table 10

*Model 8: The General Model with One Intersection Removed (TOO HARD RATIO and SCHOOL)*

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.74</td>
<td>.04</td>
<td>1</td>
<td>20.85</td>
<td>.0305</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First_Try_Ratio</td>
<td>5.83</td>
<td>.72</td>
<td>72</td>
<td>8.07</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Too_Hard_Ratio</td>
<td>-2.93</td>
<td>.12</td>
<td>72</td>
<td>-23.53</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-.09</td>
<td>.09</td>
<td>72</td>
<td>-1.05</td>
<td>.2976</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELL_Section</td>
<td>.60</td>
<td>.12</td>
<td>72</td>
<td>4.82</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td>.24</td>
<td>.03</td>
<td>72</td>
<td>8.03</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First_Try*ELL_Sectio</td>
<td>.98</td>
<td>1.37</td>
<td>72</td>
<td>.72</td>
<td>.4767</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First_Try_Rat*School</td>
<td>-4.14</td>
<td>.61</td>
<td>72</td>
<td>-6.80</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Too_Hard_*ELL_Sectio</td>
<td>.34</td>
<td>.74</td>
<td>72</td>
<td>.47</td>
<td>.642</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender*ELL_Section</td>
<td>1.05</td>
<td>.16</td>
<td>72</td>
<td>6.62</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender*School</td>
<td>-.013</td>
<td>.06</td>
<td>72</td>
<td>-1.94</td>
<td>.0558</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Covariance Parameter Estimates*

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Section</td>
<td>1.36E-20</td>
<td>.</td>
</tr>
</tbody>
</table>
Model 9: The General Model with Two Interactions Removed (FIRST TRY RATIO - ELL SECTION and TOO HARD RATIO - SCHOOL)

The Solution for Fixed Effects section shows that the value of the INTERCEPT (.01) is significantly different from zero at the .05 level. The slopes for FIRST TRY RATIO (<.0001), TOO HARD RATIO (<.0001), ELL SECTION (<.0001), and SCHOOL (<.0001), as well as the interactions between FIRST TRY RATIO and SCHOOL (<.0001), and GENDER and ELL SECTION (<.0001), were highly significant at .05 level. The interaction between TOO HARD RATIO and ELL SECTION (.91), and GENDER and SCHOOL (.07) were non-significant at the .05 level of significance, as well as the slope for GENDER. The Covariance Parameter Estimates shows that there was no unexplained variability in this model.
Table 11

Model 9: The General Model with Two Interactions Removed (FIRST TRY RATIO - ELL SECTION and TOO HARD RATIO - SCHOOL)

### Solutions for Fixed Effects

| Effect                  | Estimate | Standard Error | DF  | t Value | Pr > |t|  |
|-------------------------|----------|----------------|-----|---------|-------|   |  |
| Intercept               | 0.72     | 0.02           | 1   | 42.70   | 0.0149|   |  |
| First_Try_Ratio         | 6.24     | 0.17           | 73  | 35.75   | <.0001|   |  |
| Too_Hard_Ratio          | -2.84    | 0.10           | 73  | -27.41  | <.0001|   |  |
| Gender                  | -0.05    | 0.12           | 73  | -0.44   | 0.6624|   |  |
| ELL_Section             | 0.65     | 0.08           | 73  | 8.26    | <.0001|   |  |
| School                  | 0.25     | 0.02           | 73  | 11.63   | <.0001|   |  |
| First_Try_Rat*School    | -4.25    | 0.47           | 73  | -9.00   | <.0001|   |  |
| Too_Hard_*ELL_Sectio    | 0.11     | 1.03           | 73  | 0.11    | 0.9120|   |  |
| Gender*ELL_Section      | 0.96     | 0.24           | 73  | 4.03    | 0.0001|   |  |
| Gender*School           | -0.14    | 0.07           | 73  | -1.83   | 0.0707|   |  |

### Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Section</td>
<td>1.05E-20</td>
<td>.</td>
</tr>
</tbody>
</table>
Model 10: The General Model with Three Interactions Removed (FIRST TRY RATIO - ELL SECTION, TOO HARD RATIO - SCHOOL, and TOO HARD RATIO - ELL SECTION)

The Solution for Fixed Effects section shows little change in values of the estimates compared to the previous model. The Covariance Parameter Estimates show that there is no unexplained variability in this model. The convergence criterion was met in 18 iterations. However, the fit of the model was improved as I describe in the Model Comparison section.

Model 11: The General Model with all Interactions Removed

The Solutions for Fixed Effects section shows that the value of the Intercept was not significantly different from zero, while the other slopes for all five covariates used (3 at Level 1 and 2 at Level 2) were significantly different from zero at the 5% level. Except for the “TOO HARD” variable which correlated negatively with the success in solving the problems, all other variables correlated positively with the dependent variable. The convergence of the model was reached after 15 iterations.
Table 12

**Model 11: The General Model with all Interactions Removed**

| Effect             | Estimate | Standard Error | DF | t Value | Pr > |t| |
|--------------------|----------|----------------|----|---------|-------|---|
| Intercept          | -.23     | .19            | 1  | -1.19   | 0.44  |
| First_Try_Ratio    | 2.81     | .73            | 77 | 3.84    | 0.0003|
| Too_Hard_Ratio     | -3.29    | .27            | 77 | -11.90  | <.0001|
| Gender             | .15      | .05            | 77 | 3.00    | 0.003 |
| ELL_Section        | .68      | .28            | 77 | 2.39    | 0.01  |
| School             | .21      | .10            | 77 | 2.01    | 0.04  |

**Covariance Parameter Estimates**

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Section</td>
<td>0.00E+00</td>
<td>.</td>
</tr>
</tbody>
</table>

The following table (Table 13) shows the values of -2 Res Log Pseudo-Likelihood (-2RLPL), Generalized chi-square, and Generalized chi-square per degree of freedom (Gen Chi-Square/DF) for each of the last four models. I noticed that all Models 7-11 obtained a value of Gen Chi-Square/DF within the interval 1±.10. The closest value to 1 was achieved for Model 10 (.99), which I considered to be the best fit and therefore retained for further considerations.
Before I could use Model 10 as my final model, I needed to explore its statistical assumptions. According to Tabachnick and Fidell (2007), the residuals should meet the following assumptions: (1) they should be normally distributed around the predicted scores (normality); (2) they should be in a straight-line relationship with the predicted scores (linearity); and (3) the variance of the residuals should be the same for all predicted scores (homoscedasticity). The scatterplot tends to be skewed to one end or another when normality is violated. In addition, it is C-shaped or U-shaped when linearity is violated, or fans out toward one end when homoscedasticity is violated. When I inspected the residuals’ distribution (Figure 8) for Model 10, the scores clustered in the center (Figure 8a), had a bell-shaped distribution symmetrical about the mean (Figure 8b), were randomly scattered about a horizontal line (Figure 8c), and centrally

<table>
<thead>
<tr>
<th>Model</th>
<th>-2RLPL</th>
<th>Gen Chi-Square</th>
<th>Gen Chi-Square/DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 7</td>
<td>106.27</td>
<td>72.46</td>
<td>1.03</td>
</tr>
<tr>
<td>Model 8</td>
<td>109.16</td>
<td>92.56</td>
<td>.97</td>
</tr>
<tr>
<td>Model 9</td>
<td>114.22</td>
<td>72.70</td>
<td>.98</td>
</tr>
<tr>
<td>Model 10</td>
<td>118.78</td>
<td>72.72</td>
<td>.99</td>
</tr>
<tr>
<td>Model 11</td>
<td>130.14</td>
<td>80.38</td>
<td>1.03</td>
</tr>
</tbody>
</table>
concentrated and distributed in a rectangular pattern (Figure 8d). Thus, the assumption of normality, linearity and homoscedasticity for residuals were met, and I concluded that Model 10 was reliable with normally distributed random coefficients.

**Figure 8**. The normality of residuals for the Model 10. Each panel consists of: (a) a scatterplot of the residuals; (b) a histogram with normal density; (c) a Q-Q plot; and (d) a box plot of the residuals.

As with all multilevel models, variability in the level-1 slopes represents the between-groups variability in the relationship between level-1 predictors and the success ratio outcome. Table 14 shows that the intercept for the final model was $\gamma_{00} = 0.72$ (p=.01).
Table 14

Model 10: The Final Model

| Effect                     | Estimate | Standard Error | DF | t Value | Pr > |t| |
|----------------------------|----------|----------------|----|---------|-------|-----|
| Intercept                  | .72      | .02            | 1  | 43.86   | .0145 |     |
| First_Try_Ratio           | 6.23     | .12            | 74 | 51.08   | <.0001|     |
| Too_Hard_Ratio            | -2.81    | .29            | 74 | -9.84   | <.0001|     |
| Gender                    | -.06     | .15            | 74 | -.38    | .7030 |     |
| ELL_Section                | .65      | .07            | 74 | 9.82    | <.0001|     |
| School                    | .25      | .02            | 74 | 11.02   | <.0001|     |
| First_Try_Rat*School      | -4.24    | .41            | 74 | -10.27  | <.0001|     |
| Gender*ELL_Section        | .97      | .31            | 74 | 3.11    | .0027 |     |
| Gender*School             | -.14     | .07            | 74 | -1.83   | .0707 |     |

Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Section</td>
<td>2.81E-20</td>
<td></td>
</tr>
</tbody>
</table>

This value of the intercept represents the expected log odds for an EP female student (GENDER=0, ELL SECTION=0) from the AMS (SCHOOL=0), who had 45% of the problems solved correctly from the first attempt (FIRST TRY RATIO=0) and considered 10% of all problems to be too hard (TOO HARD RATIO=0). The values of .45 and .1 for the FIRST TRY RATIO and TOO HARD RATIO variables represent their original mean values. Prior to running Model 10, I centered these variables so that the
new mean values were zero in order to interpret the intercept. Thus, the estimated odds for a student with these individual and school characteristics were $\exp (.72)= 2.05$, corresponding to a predicted probability of success of .67 (from equation 1.10). The value 2.05 is called the referent odds, and represents the prediction of the odds in the case when all predictors (school- and child-level) are zero. According to these referent odds, there was a 67% chance for an EP female student from the AMS to be successful in solving word problems when she solved 45% of the problems correctly on the first attempt and considered 10% of the problems to be too hard.

The estimated odds, and thus the predicted probability of success, tended to increase for all sections from LCMS with the number of problems solved correctly on the first attempt. Surprisingly, the predicted probability of success also increased as the rating of the word problems considered too hard increased. Specifically, the effect of ELL SECTION on the log odds was positive and statistically different from zero, $\gamma_{01} = .65$ ($p<.001$), when controlling for FIRST TRY RATIO, TOO HARD RATIO, GENDER and SCHOOL. This effect suggested that when FIRST TRY RATIO, TOO HARD RATIO, GENDER, and SCHOOL were held constant, the odds of success was expected to increase by 1.91 ($\exp (.65)=1.91$) corresponding to a predicted probability of success of .65 (from equation 1.10) as the ELL SECTION increased by one unit. A one-unit change indicates a difference of one standard deviation in each measure. The within-unit variance explained is a measure of how well the independent variables in the model explain the outcome variable. The between-unit measure is the amount of variance between level-2 units that is accounted for by the predictors in the model.
From Table 2, the ELL SECTION has a standard deviation of .13. Therefore, one standard deviation difference in ELL SECTION is associated with the difference in log odds of success of .13*.65=.08 or a relative odds of \( \exp(.08) = 1.08 \) corresponding to a predicted probability of 51% (from equation 1.10). Two similar classes/sections that differ by one standard deviation in ELL SECTION could be expected to be .08 units apart in log odds of success corresponding to a predicted probability of 51%. For example, an EP female student from AMS who solved 45% of the problems correctly on the first attempt and considered 10% of the problems to be too hard, had a 67% chance to be successful in solving word problems. Adding one unit to the ELL SECTION variable would lead to a predicted log odds of .72+.65 = 1.37, associated with a predicted probability of success of 79%. An additional unit increase in ELL SECTION variable would lead to a predicted log odds of .72+2*.65=2.02, corresponding to a predicted probability of success of 88%.

The effect of FIRST TRY RATIO on success holding TOO HARD RATIO, SCHOOL, GENDER, and ELL SECTION constant, was positive and statistically significant, \( \gamma_{10} = 6.23 \) (\( p<.001 \)). This effect suggested that the expected odds of success for students who solved problems correctly on the first attempt were \( \exp(6.23) = 507.75 \) times the odds of success of an other-similar student who had 0% chance of success in solving problems correctly from the first attempt. With each one unit increase in the FIRST TRY RATIO, a child’s odds of reaching success increased by 99% (from equation 1.10), holding other effects constant.

The effect of TOO HARD RATIO on success, holding FIRST TRY RATIO, SCHOOL, GENDER, and ELL SECTION constant, was negative and statistically
significant, $\gamma_{20} = -2.80$ (p<.001), equivalent with the odds ratio of $\exp(-2.80)=.06$. This effect suggested that with each one unit increase in the TOO HARD RATIO, a child’s odds of reaching success decreased by 94% (that is, $100\% \times (.06-1)= - 94\%$), holding other effects constant.

The effect of GENDER on success, holding TOO HARD RATIO, FIRST TRY RATIO, SCHOOL, and ELL SECTION constant, was not statistically significant at the .05 level of significance. However, I did not eliminate the variable Gender because of its statistical significance in the interaction with the ELL SECTION.

The results of the final model from the Covariance Parameter Estimates (Table 14) showed that there was not residual variability in the group intercepts ($\tau_{00}=0$). The absence of variability confirms that the model has enough child- or group-level predictors.

To determine if any additional insight could be gained, I also conducted post hoc data analysis on the final model. In this analysis I decomposed the significant interactions so that each path became its own model. For example, the interaction between FIRST TRY RATIO and SCHOOL was broken down into low score for FIRST TRY RATIO, high score for FIRST TRY RATIO, SCHOOL 0 and SCHOOL 1. I re-ran the four model components individually using only one predictor at a time. I used the same method to analyze the interaction between GENDER and ELL SECTION. Results showed that all four paths for the first and second interaction remained statistically non significant predictors of log odds of success when using only one predictor at a time. Although all the interactions were significant, the final model did not capture enough
information to explain the interaction effects between FIRST TRY RATIO and SCHOOL, and between GENDER and ELL SECTION.

CHAPTER 5
CONCLUSIONS

In this chapter I present the conclusions drawn from my study. First, I summarize the study. Second, I discuss the research topics, using the literature review and my research findings. Third, I discuss the limitations of the study. Finally, I present the importance of the study and make recommendations for further research.

Summary of the Study

I analyzed the mathematics performance of 84 seventh-grade students attending two public middle schools from the same school district in the Southwest. My goal was to determine the students’ abilities in solving algebra word problems in a web-based environment. The software, called Animal Watch, was designed to help students learn mathematics through interactive display. I conducted this analysis using a multilevel-modeling approach. I considered the variables related to the students at level 1, and those related to the four classes in which students have been clustered at level 2. Out of the variables measured in the experiment, I found that only five variables had a significant impact upon the success ratio of the attempted problems: FIRST TRY RATIO (number of problems solved correctly on the first attempt over the number of problems solved correctly), TOO HARD RATIO (number of problems considered to be too hard over the number of problems attempted), ELL SECTION (number of ELs in each class over the total number of students of that class), SCHOOL, and GENDER.
Research Topic One

In the first research topic I investigated if the ELs performed better, worse or the same as the EPs. As research shows, the issue of creating equal opportunities in mathematics education for ELs is not completely understood (Cuevas, 1984; Enyedy, Rubel, Castellon, Mukhopadhyay, Esmonde & Secada, 2008; Moschkovitch, 2002, 2005). Learning English in an instructional context is considerably more demanding than acquiring basic conversational proficiency. In the specific case of mathematics learning, students must understand the mathematics register, as well as the mathematics-specific meanings for words that may be familiar in other contexts.

Moreover, different studies discuss the concerns related to the correct classification of students with respect to their English level in mathematics learning and testing (Echevarria, Powers & Short, 2006; Sireci & Khaliq, 2002; Solano-Flores, 2008). While various tests exist for determining student English proficiency in general, they may be inefficient in finding the true dynamics in a mathematics class where examples are given in terms of sports or other activities that might be unknown to the students. When the new concepts are based on an unknown word (like inning from the game of baseball) the consequences could be major, even if the student might otherwise show a relatively good understanding of English vocabulary. According to Cummings’ hypothesis, bilingual students who are highly proficient in both languages usually have better results
than their monolingual peers, yet no one has been able to establish a methodology to determine the efficiency of the threshold.

In my study, I established the students’ linguistic abilities in English and their primary language based on previous testing and their SES. Overall, I concluded that the ELs in the study were proficient in both English and their native language.

With regard to English language knowledge, I used four language proficiency tests to estimate the ELs’ level: Arizona English Language Learner Assessment (AZELLA), new Arizona English Language Learner Assessment AZELLA 2 Form AZ-2 (AZ2), Stanford English Language Proficiency (SELP), and IDEA Oral Language Proficiency Test (IPT). From a total of 26 ELs the district classified four as having an intermediate English status and the other 22 as having a proficient English status. No students with a low English status were identified.

With regard to native language knowledge, I used social economic status (SES) and enrollment in special education to classify the ELs’ native language skills (Pray, 2005). I used students’ enrollment in the lunch program to determine SES. Those students who received a free or reduced lunch were considered “middle level in native language” and those who paid full price for lunches were considered “upper level in native language.” Those students who were enrolled in a special education program were considered “low level in native language.” As a result, of 26 ELs, two students were classified as having a low level in their native language, four students as having a middle level, and the rest as having an upper level.

In addition, I found that the students from the classes that had a higher percentage of ELs performed better than those in the classes where the EL concentration was lower.
The results of the statistical analysis showed that the EL concentration in each class (ELL SECTION) was statistically significant at a level $p < .001$ in explaining the individual performance (log-odds of success). This result is significant because the EL status considered as a variable at Level 1 did not make an important contribution to the rate of success. In other words, what ensured a better rate of success in solving the word problems was not being an EL student per se, but being part of one of the classes with a higher percentage of ELs.

According to Cummins’ Threshold hypothesis, ELs are expected to perform better than EPs once they have exceeded a specific threshold in English and in their primary language. The ELs in my study had achieved this threshold. If they had not, then being an EL student in a class with higher ELs concentration would have resulted in poorer results. Unfortunately, the methodology I employed did not allow me to identify this Threshold as an independently measurable indicator.

In addition, I could use Cummins’ Threshold hypothesis only in a limited manner, to study EL groups among classes, but not individual ELs. The EL status variable considered at Level 1 did not make any significant contribution to the results, but I was able to determine the differences among classes with a different number of ELs. Since this study did not include class observations, I did not have any prima facie evidence about possible differences in teaching styles or class dynamics, elements which could provide a reasonable explanation for this apparent discrepancy. Moreover, Cummins’ Threshold hypothesis suggests that the English language abilities of the ELs in my study aided them. In particular, the ELs were able to solve the word problems in the test and were not disadvantaged in learning mathematics compared to EPs.
In addition, I found only one significant interaction between ELL SECTION and GENDER, which I discuss in the next Research Topic.

**Research Topic Two**

In this topic I investigated the differences in mathematics abilities between boys and girls and the nature of this relationship. According to some researchers (Seegers & Boekaerts, 1996; Hyde 1990), boys are expected to show a slight superiority in mathematics skills when compared to girls, with this difference increasing with age. More specifically, boys become better than girls in performing mathematics tasks involving more complex abstract manipulation. On the other hand, girls are slightly better than boys in performing standard computations learned in elementary and middle school. McGraw et al. (2006) determined that the slight gap favoring boys has remained constant over the last two decades, and its magnitude was affected by race, SES, and students’ placement on an aptitude scale (the gap between higher performing boys and girls was higher that the gap between the lower performers).

Motivation has also been shown to be well differentiated between genders, with boys more confident in their mathematics abilities, even when their results were comparable with those of girls who showed more willingness to learn and to put in hard work (Fennema 2000). A key reason boys are more confident is that teachers consider them to be better in mathematics than girls (Brophy & Good, 1970; de Boer et al., 2010).

In my study boys and girls had to solve problems that contained both basic computations (addition, subtraction, multiplication, and division) as well as abstract thinking at a higher level. Since the problems represented a mixture of basic computation
and higher abstract thinking, I could not formulate any a priori expectations about gender success.

The results showed that gender by itself did not represent a significant indicator of success rate. In my research model, the p-value for gender was \( p = .703 \), a value far from the standard 5% required for significance. Yet, interestingly, I found a significant interaction between gender and the percentage of ELs in each class (GENDER*ELL_SECTION, \( p = .0027 \)). The positive sign of this interaction indicated that the class performed better when there were more EL male students. In other words, the overall performance of classes with more EL male students was expected to exceed those of classes with more EL female students, as well as those of classes with more EP male students and classes with more EP female students.

The magnitude of this difference can be illustrated by the following simple thought experiment. Assume that the same test is carried out on two different sets of students with absolutely identical characteristics as the ones considered here, but in one case all students were EL males and in another one all students were EP females. Then, the expected difference in log-odds of success would be .97 in favor of the classes selected from the former group. That is equivalent to saying that the group made up of only EL male students would have 72.5% more chances of solving the test correctly than the group made up of only EP females. If this point estimate looks rather large, a better indicator would be to calculate the 99% confidence interval CI_{99\%} = [.97-2.58*.31, .97+2.58*.31] or CI_{99\%} = [.17, 1.77]. Then, the expected difference in probabilities between the two groups is found to lie—with 99% probability—with the interval [.54, .85].
In the absence of additional data, I can not easily formulate a definitive explanation for this interaction. The most plausible justification could be obtained by corroborating the Cummins’ Threshold hypothesis and McGraw et al.’s findings that boys slightly outperform girls in mathematics abilities. Moreover, it is conceivable that additional variables not measured in my study, such as teaching style and class interactions (Mulis et al., 2000; Stipek & Granlinski, 1991; Brohpy & Good 1970), could act as covariates capable of influencing this result.

Finally, I found a non-significant interaction between gender and school \((p=.0707)\) despite the fact that the SCHOOL by itself represented a valid indicator at Level 1. This could indicate that both schools had a similar rate of success in teaching mathematics to boys when compared to girls. Since the GENDER variable in itself was not a significant indicator, I concluded that, in each school, boys and girls learned the same amount of mathematics.

My study departs from the published research. Whereas previous studies demonstrated that girls do better on computational problems, and boys do better on problems requiring higher reasoning processes, I discovered that classes with more EL males learn mathematics better than classes with more EP females.

**Research Topic Three**

In my third research topic I investigated if some of the students’ behaviors during the test could be valid indicators for overall test success. To the best of my knowledge, such an investigation has not been carried out previously, as the computer-based mathematics learning environment is still a novelty.
As I mentioned in the description of my study, I tested multiple variables when students solved the problems:

- success in solving the problem on first, second or third attempt (FIRST TRY, SECOND TRY, THIRD TRY, NUMBER ATTEMPTS, STRIKEOUT),
- initial decision made if the problem was too hard (TOO HARD),
- time spent per problem (TIME),
- if the student sought help (USED HELP).

Out of all these variables, I determined there were only two valid indicators at the student level (Level 1): FIRST_TRY_RATIO and TOO_HARD_RATIO. The significance level for both variables was $p < .0001$, and their slopes had opposite signs. The slope value for FIRST_TRY_RATIO was $+6.28$, which showed a positive correlation between the ability to solve the problems on the first attempt and the success ratio in solving all problems. This positive correlation indicated that the better students were capable of finding the correct answer faster and with less guess work. In contrast, the percentage of problems solved from the second and third attempts did not correlate significantly with the ratio of problems solved correctly. This indicates that students who did not solve the problems on the first attempt could not really benefit from the additional attempts, even when help was available.

The second significant indicator, TOO_HARD_RATIO, had a slope with a negative value, which indicated a negative correlation between the percentage of problems solved correctly and those considered too hard from the very beginning. However, the limitations of my study design did not allow me to establish if this was simply the result of the students having attempted fewer problems, or the result of
motivational factors (such as reduced confidence in their abilities or a diminished motivation for the test itself).

**Limitations of the Study**

A major limitation of my study was the reduced sample size. Participants came from only two schools from the same school district, and from only one grade level (seventh). This limitation was acceptable, however, because the district has a diverse student population, and the schools were randomly selected.

Statistical assumptions were another limitation. Multilevel linear modeling is an extension of multiple linear regressions, and the limitations and assumptions of the linear regression apply to all levels of analysis. One of the assumptions of regression analysis is that the independent variables are measured without error, a clear impossibility in most social and behavioral science research. The violation of this could lead to a poor fit of regression models, and therefore to unreal results.

Moreover, regression solution is extremely sensitive to the combination of variables that are included within it. Regression analyses reveal the relationship among variables but not the causality. Demonstration of causality is a theoretical, logical, and experimental, rather than statistical, problem. An apparently strong relationship between variables could stem from many sources, including the influence of unmeasured variables. For this reason, choosing the independent variables should be based on theory, observations, and careful examination of the distribution of residuals—something that always limits research.

A final limitation of my study is its narrow area of investigation. The students were tested only in algebra word problems, and these problems did not cover a large area
of the mathematics standards. To obtain meaningful correlation with students’ previous achievements, the tests need to be modeled after the Common Core State Standards and the Mathematics Standards articulated by grade level.

Even though the study does have some limitations, its findings and conclusions are still valuable. It provides a picture of how a web-based tutoring system for solving algebra word problems impacts diverse students. Such a study has not been examined in the literature to date.

**Importance of the Study and Recommendations for Further Research**

Web-based environments for mathematics learning and testing are in their infancy, and the number of studies dedicated to their efficiency is limited. From this perspective, my study represents an attempt to better understand this type of instruction. The MLM statistical tools employed in the study allowed me to unravel some of the finer details and interactions between various indicators and variables measured. In addition, the GLMMIX procedure from SAS that I used in this study has been rarely employed in educational studies. Although it is considered a powerful instrument for this kind of investigations, the vast majority of researchers in the field still rely upon the older GLM routine for which the normality of the DV requirements is rarely truly satisfied.

Of the results discussed above, two are of real interest and warrant further investigation, because they have never been reported before. First, the positive correlation between the rate of success and the number of ELs in each class suggested that in the case of mixed EP-EL classes, a critical mass of ELs might be necessary for mathematics instruction to be beneficial for them. Thus, further investigations should
concentrate on: (a) confirming my conclusion, (b) determining a value for this threshold, and (c) establishing if my conclusion is a result of a teaching methodology triggered by the larger number of ELs in the class, or that of better collaboration among students.

Second, the interaction between the percentage of ELs in each class and their gender suggested a specific EPs/ELs/girls/boys student distribution in mathematics classes. As my study revealed, the mathematics performance for classes with more EL males was superior to classes with more EP students or EL girls. This result, if confirmed by subsequent research, could suggest a way for optimal EPs/ELs/girls/boys student distribution in mathematics classes. Due to the design and limited number of students in my study, I was not able to determine clear boundaries for such a distribution.

In addition, as computers and web-learning environments are used more and more frequently in schools, their impact will likely continue at an increased pace. However, there is no clear understanding of what the advantages and disadvantages of computer- and web-based instruction are, and no coherent guidelines for how it should be implemented.
APPENDIX A

CONSENT FORM FOR STUDENTS
MINOR'S ASSENT FORM

Your mother/father has told me it was okay for you to participate in this study. You will work one session of no more than one class period with a software in your classrooms or school computer lab under the direction of your teacher. The instructional modules included in the computer program will include a tutorial in English mathematics vocabulary (e.g., “dividend” “tangent” “mixed number”) that may not be familiar to English Language Learners, a feature through which teachers can choose word problems with low grade-level readability for the ELLs in their classes, and interactive support for forming equations from word problem text on the computer screen. The purpose of this project is to develop a computer program to help students learn to solve mathematics word problems. When the program is ready, it will be available for free to students in states with large numbers of English Language Learners.

Your participation in this study is voluntary. You may decide to not begin or to stop the study at any time. Your refusing to participate will have no effect on your student status.

______________________________
Child's Name

______________________________
Child’s Signature     Date

______________________________
Presenter’s Signature    Date

Otilia Barbu   3/10/2010

______________________________
Investigator's Signature     Date
APPENDIX B

CONSENT FORM FOR PARENT/ GUARDIAN
PARENT/LEGAL GUARDIAN PERMISSION FORM

You are being asked to read this form so that you know about this research study. Federal regulations require that you know about the study and the risks involved. If you decide that you want your child to participate in this study, signing this form will say that you know about this study. If you decide you do not want your child to participate, that is okay. There will be no penalty to you or your child and your child will not lose any benefit which s/he would normally have.

WHY IS THIS STUDY BEING DONE?
You have the choice for your child to participate in this project.

The purpose of this project is to develop a computer program to help students learn to solve mathematics word problems. Mathematics word problems can be especially hard for students who are learning to read English. We want to provide tutoring to English Language Learners to find out what is most helpful for them. We will use the results to guide the design of our computer program. When the program is ready, it will be available for free to students in states with large numbers of English Language Learners.

WHY IS YOUR CHILD BEING ASKED TO BE IN THIS STUDY?
To be in this study, your child must be an English Language Learner enrolled in Grade 6 or 7.

WHAT ARE THE ALTERNATIVES TO BEING IN THIS STUDY?
The alternative is that your child does not have to participate in the study.

WHAT WILL YOUR CHILD BE ASKED TO DO IN THIS STUDY?
Your child’s participation in this study will include one session of no more than one class period under the direction of the teacher. The instructional modules included in the computer program will include a tutorial in English mathematics vocabulary (e.g., “dividend” “tangent” “mixed number”) that may not be familiar to English Language Learners, a feature through which teachers can choose word problems with low grade-level readability for the ELLs in their classes, and interactive support for forming equations from word problem text on the computer screen. Students will work with the mathematics tutoring software as part of their mathematics class learning activities, under the direction of their classroom teacher. The purpose of this project is to develop a computer program to help students learn to solve mathematics word problems.

ARE THERE ANY RISKS TO MY CHILD?
Students sometimes get frustrated if they cannot solve a mathematics problem. However, the software will help the student learn to solve each problem.

ARE THERE ANY BENEFITS TO MY CHILD?
We hope that your child may learn some useful strategies for solving word problems in English by working with the software.

**WILL INFORMATION FROM THIS STUDY BE KEPT CONFIDENTIAL?**
We do not need to record any identifying information about your child. The data collected are entirely electronic and consist of flat files representing a user’s actions on a series of word problems. Users are assigned user accounts (e.g., AJHS4237) by teachers, who will indicate to researchers which accounts were assigned to ELLs. No information that could be linked to individual students is recorded or retained. People who have access to your child’s information include the Principal Investigator and study personnel. In addition, representatives of regulatory agencies (including the Human Subjects Protection Program) may access your child’s records to make sure the study is being run correctly and that information is collected properly. The agency that funds this study (sponsor) may also see your child’s information. However, any information they have will be coded with a number so that they cannot tell who your child is. If there are any reports about this study, your child’s name will not be in them. This consent form will be filed in an official area.

**WILL I OR MY CHILD BE PAID TO BE IN THIS STUDY?**
There is no cost to you or your child for being in this study except your child’s time. You/your child will not be paid for being in this study.

**WHO CAN BE CONTACTED ABOUT THIS STUDY?**
You or your child can call the Principal Investigator to ask questions or tell her about a concern or complaint about this research study. The Principal Investigator is Otilia Barbu, and you can call her on her cell phone at (520) 975-7694. If you have questions about your child’s rights as a research subject you or your child may call the Human Subjects Protection Program office at (520) 626-6721. If you or your child have questions, complaints, or concerns about the research and cannot reach the Principal Investigator; or want to talk to someone other than the Investigator, you or your child may call the Human Subjects Protection Program office. (If out of state use the toll-free number 1-866-278-1455.)

**STATEMENT OF CONSENT**
The procedures, risks, and benefits of this study have been told to me and I agree for my child to be in this study and sign this form. My questions have been answered. I may ask more questions whenever I want. My child can stop participating in this study at any time and there will be no bad feelings. My child’s medical care will not change if s/he quits. The researcher can remove my child from the study at any time and tell me why my child has to stop. New information about this research study will be given to me/my child as it is available. My child and I do not give up any legal rights by signing this form. A copy of this signed consent form will be given to me.

____________________________________  __________________________________
Subject's Name

___________________________________  ______________________________

Parent/Legal Guardian

___________________________________  ______________________________

Parent/Legal Guardian

___________________________________  ______________________________

INVESTIGATOR'S AFFIDAVIT:
Either I have or my agent has carefully explained to the subject the nature of the above project. I hereby certify that to the best of my knowledge the person who signed this consent form was informed of the nature, demands, benefits, and risks involved in his/her participation.

___________________________________  ______________________________

Signature of Presenter

Signature of Investigator

Please sign one copy and return it to your child’s mathematics teacher. You can keep one copy so that you have all the information in case you have questions later on.
APPENDIX C

SITE AUTHORIZATION LETTER
Dear Ms. Barbu:

The District Research Committee has completed a review of your research proposal. The action research proposal for “Mathematics Word Problems Solving by English Language Learners and Web Based Tutoring System” has been approved.

Please contact the Principal of the school before proceeding with your research. If you have any questions, feel free to contact me at 696-5172.

Sincerely,

Patrick Nelson
Associate Superintendent
APPENDIX D

WORD PROBLEMS
**Introductory Level (EM-EE)**

Many zoos around the world are trying to raise Wild Horses. One zoo had 11 Wild Horses. By the end of the year, 3 baby horses had been born. How many wild horses did the zoo have then?

Word count: 38
REAP readability: Grade 3

**Intermediate Level**

a) EM-HE

The Snow Leopard has an unusually long tail, which it often uses to shield its face for warmth. Its tail and body are equivalent in length. If its tail is 2 feet long, what is the total length of the Snow Leopard?

Word count: 42
REAP readability: Grade 9

b) HM-EE

Many people think that Pandas are cute. But Pandas are big and can be dangerous. An adult male Panda can weigh 123 kilograms. One kilogram equals 2.2 pounds. What is the Panda’s weight in pounds?

Word count: 35
REAP readability: Grade 1

**Challenging Level (HM-HE)**

The Condor is an enormous bird with powerful wings that allow it to soar to extraordinary heights. One Condor was documented flying at an altitude of 14,750 feet. If one foot is equivalent of 12 inches, what was the Condor’s altitude in inches?

Word count: 43
REAP readability: Grade 12
REFERENCES


Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and


