2008

# Correlates of mathematics achievement in developed and developing countries: An HLM analysis of TIMSS 2003 eighth-grade mathematics scores

Ha T. Phan
*University of South Florida*

# Correlates of Mathematics Achievement in Developed and Developing Countries:

# An HLM Analysis of TIMSS 2003 Eighth-Grade Mathematics Scores

by

Ha T. Phan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Educational Measurement and Evaluation
College of Education
University of South Florida

Major Professor: Jeffrey D. Kromrey, Ph.D.
Robert F. Dedrick, Ph.D.
John M. Ferron, Ph.D.
Christina Sentovich, Ph.D.

Date of Approval:
October 10, 2008

Keywords: secondary data, math performance, multilevel analysis, large-scale assessment, international research

# DEDICATION

This manuscript is dearly dedicated to my family. To my father and mother whose positive influences on my life, my education, and the work presented here, have been tremendous. To my wonderful husband, Son and my two beautiful children, Thao and Namson, whose endless love and caring and countless support have made my educational dream possible.

# ACKNOWLEDGEMENTS

I would like to thank the following individuals for their significant contributions toward my successful completion of this degree. First, I would like to thank Jeffrey Kromrey, my major professor, for his wonderful guidance, support, and wisdom that benefited me in many ways throughout this dissertation journey. I am also in gratitude to my committee, John Ferron, Robert Dedrick, and Christina Sentovich for their insightful feedback and efforts in the preparation of this manuscript. I am grateful to Gladis Kersaint and Denisse Thompson from the Math Education Program for their valuable consultations and recommendations related to the work presented here. Also, I am grateful to Lou Carey and Cynthia Parshall who first inspired my interest in measurement and research. Finally, I would like to acknowledge the National Center for Education Statistics for their training in large-scale data analysis as well as their provision of the data for this research.

TABLE OF CONTENTS

## LIST OF TABLES

LIST OF FIGURES

Correlates of Mathematics Achievement in Developed and Developing Countries:
An HLM Analysis of TIMSS 2003 Eighth-Grade Mathematics Scores

Ha T. Phan

ABSTRACT

Using eighth-grade mathematics scores from TIMSS 2003, a large-scale international achievement assessment database, this study investigated correlates of math achievement in two developed countries, Canada and the United States and two developing countries, Egypt and South Africa. Variation in math achievement within and between schools for individual countries was accounted for by a series of two-level HLM models. Specifically, there were five sets of HLM models representing student background, home resources, instructional practices, teacher background, and school background related factors. In addition, a final model was built by including all the statistically significant predictors in earlier models to predict math achievement. Findings from this study suggested that whereas the instructional practices model worked the best for the United States and the teacher background model served as the most efficient and parsimonious model for predicting math achievement in Egypt, the final model served as the best model for predicting math achievement in Canada and South Africa. These findings provide empirical evidence that different models are needed to account for factors related to achievement in different countries. This study, therefore, highlights the importance that policy makers and educators from developing countries should not base their educational decisions and educational reform projects solely on research findings of

developed countries. Rather, they need to use their country-specific findings to support their educational decisions. This study also provides a methodological framework for applied researchers to evaluate the effects of background and contextual factors on students' math achievement.

# CHAPTER ONE

# INTRODUCTION

## Statement of Problem

Students' mathematics achievement is often associated with the future economic power of a country (Baker & LeTendre, 2005; Bush, 2001; Heyneman & Loxley, 1982, 1983; Wobmann, 2003). Thus, the desire to understand and identify factors that may have meaningful and consistent relationships with math achievement has been commonly shared among national leaders and policy makers as well as educators around the world. For example, in 2007, there were more than 60 countries participating in the Trends in International Mathematics and Science Study (TIMSS) (TIMSS, 2007). By collaboratively supporting and participating in a large-scale international achievement study such as the TIMSS, it was hoped that the rich data (achievement and other contextual data) collected from such a study could illuminate important correlates of math achievement both within and between countries that would "otherwise escape detection" (Wagemaker, 2003, p.1).

Unfortunately, despite the fact that data from these international achievement studies have been made publicly available for all participating countries [National Center for Education Statistics (NCES), 2007] only a small number of these countries was included in subsequent research studies. A review of existing literature suggested that low income countries as well as those that performed poorly in international achievement studies such as South Africa, Chile, and Egypt were rarely included in international research studies (details of these studies are provided in Chapter Two). In contrast, researchers tended to focus on a small group of developed and high-performing countries such as Japan, Korea, Hong Kong, Singapore, Germany, Canada, and the United States. Such bias in international achievement research resulted in recent research findings related to students' math achievement that were based mostly on students in developed countries and lacking representation from developing countries. As a consequence, the problem of lacking research findings related to students' math achievement in developing countries has led many of these countries to base their educational policy decisions or even to implement educational reform projects on research findings and educational models of other developed countries (Riddell, 1997). Such bases were problematic because countries differ in characteristics and a model that worked in a developed country might not work in a developing country (Bryan et al., 2007; Delaney, 2000; Watkins & Biggs, 2001).

Given the current problem, it is very important for research studies related to international achievement to include a more diverse sample of countries (i.e., both developing and developed countries) and to utilize analytic models that yield country-specific research findings. In doing so, policy makers and educators from the developing countries that were included in these studies can use the research findings pertaining to their own countries to support their educational decisions.

## Purpose of the Study

The purpose of this study was to investigate correlates of math achievement in both developed and developing countries. Specifically, two developed countries and two developing countries that participated in the TIMSS 2003 eighth-grade math assessment were selected for this study. For each country, a series of two-level models was constructed using background and contextual factors at both the student and the classroom/teacher/school levels to account for the variance in eighth-grade students' math achievement within and between schools. Ultimately, this study aimed to produce country-specific research findings related to eighth-grade students' math achievement that can be used directly by national leaders and policy makers as well as educators from these countries, especially developing countries, to support their educational decisions. Finally, by visually and descriptively examining patterns of relationships between eighth-grade math achievement and contextual factors, this study hoped to identify important trends of relationships that tended to exist among developed and developing countries, as well as differences between these groups.

## Research Questions

The study aimed to address the following set of research questions:

1) To what extent are student background variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, and tutoring in math) associated with TIMSS 2003 eighth-grade math scores in each country?
2) To what extent are home resources variables (i.e., availability of calculator, computer, and desk for student use) associated with TIMSS 2003 eighth-grade math scores in each country?
3) To what extent are instructional variables (i.e., opportunity to learn, activities in math lessons, amount of homework assignment, and instructional time) associated with TIMSS 2003 eighth-grade math scores in each country?

4) To what extent are teacher-related variables (i.e., preparation to teach, ready to teach, and professional development) associated with TIMSS 2003 eighth-grade math scores in each country?

5) To what extent are school-related variables (i.e., class size, school resources for math instruction, and math instructional limitation) associated with TIMSS 2003 eighth-grade math scores in each country?

## Rationale for the Study

Several factors influenced the decision to use the TIMSS 2003 eighth-grade math data from four countries to investigate the relationships between student math achievement and contextual as well as background factors across countries in this study. First, mathematics was the subject of choice because of the increasingly national and international focus on math education (Baker & LeTendre, 2005; Bush, 2001; Heyneman & Loxley, 1982, 1983; Wobmann, 2003). In the United States, the topic of how to improve student achievement in math has been hotly debated for decades. The No Child Left Behind Act issued by President George W. Bush in 2001 [No Child Left Behind (NCLB) 2001] was one of the many examples that highlighted the importance of math as a school subject that has attracted attention from the country's top national leader.

Second, at the time this study was conducted, the TIMSS 2003 database provided the largest and most ambitious set of international achievement as well as background data related to students' math and science achievement at fourth-grade, eighth-grade and twelfth-grade (TIMSS, 2003). In addition, the influence of previous TIMSS findings on international education, including the U.S. education, has been widely acknowledged. As indicated in the Benchmarking Introduction of the Trends in

International Math and Science Study (TIMSS, 1999), "TIMSS results, which were first reported in 1996, have stirred debate, spurred reform efforts, and provided important information to educators and decision makers around the world" (p.16). In fact, the United States Department of Education (USDOE), through its federal entity, the National Center for Education Statistics (NCES), strongly encouraged educational researchers to use large-scale existing secondary data such as the TIMSS 2003 for research because these data "provide consistent, reliable, complete, and accurate indicators of education status and trends" (Stigler et al., 1999, p.1).

Third, eighth-grade math data were selected because of the importance of the transition period from elementary school to high school, where curriculum differentiation in math knowledge and skills is the greatest (Rodriguez, 2004). Also, as Reynolds (1991) asserted, middle-school years are a critical period for students in terms of learning math. How well students perform in math during their middle-school years is likely to determine their choices and enrollment in high school math courses. This is because courses in math are often sequential and therefore, access to advanced math courses in high school are dependent on students' success at lower level math courses at middle school (Singh, Grandville, & Dika, 2002). For example, if a student performs poorly in algebra at eighth-grade, he/she is much less likely to enroll in various math courses offered in high school. As a result, these curricular opportunities and choices further influence students' decision to enter mathematics-related fields of study at postsecondary and occupational levels. Thus, eighth-grade is an important time point to study the complex interaction of contextual factors that are potentially related to students' math achievement.

Last but not least, four countries (two developing countries and two developed countries) were sufficient for this study because this selection allowed the study to be conducted within a reasonable amount of time and resources while satisfying the important inclusion criterion of sufficient sample size and representative samples from each category of country (i.e., developing and developed countries).

## Theoretical Framework

This research study was guided by the theoretical framework of Carroll's (1963) Model of School Learning which was proposed to explain why students succeed or fail in their learning at school (Carroll, 1963). The model postulated five important factors that were theoretically related to students' success in learning: (1) Aptitude – the amount of time needed to learn the task under optimal instructional conditions, (2) Ability to understand instruction, (3) Perseverance – the amount of time the learner is willing to engage actively in learning, (4) Opportunity to learn – time allowed for learning, and (5) Quality of instruction – the extent to which instruction is presented so that no additional time is required for mastery beyond that required in regard to aptitude. Of these five factors, aptitude, ability to understand instruction, and perseverance are related to the students; whereas opportunity to learn and quality of instruction are concerned with external conditions. A thorough examination of Caroll's (1963) model of school learning and a comprehensive review of existing literature related to each of the five factors of the model are presented in Chapter Two.

## Limitations

The potential threats to the internal and external validity of this study are present at various research stages: instrument development, data collection, data analysis, and data interpretation. At the instrument development stage, the process of test adaptation and translation from the source language (i.e., English) to other target languages of the test could have made the assessment unintentionally harder or easier by translators (Hambleton, Merenda, & Spielberger, 2005). As with test item format, Wang (2001) raised a concern that students from some countries might be more familiar with test items in the constructed response format; whereas their peers in other countries might be more familiar with test items in the multiple choice format. Thus, such instrument-related issues could negatively influence the fairness of student mathematics test scores.

At the data collection stage, any discrepancies in the process of collecting TIMSS 2003 data across countries could affect the validity of the data. Because TIMSS 2003 collected data on a large-scale (i.e., in 48 countries), from multiple sources (i.e., from students, teachers, and school principals), and on different time schedules (i.e., in October and November for Southern Hemisphere countries and in April, May and June for Northern Hemisphere countries) the process of monitoring data quality could be challenging. In addition, due to country differences, some countries opted not to administer certain test or questionnaire items to their participants, resulting in some countries having no data for a set of variables (TIMSS, 2003).

At the data analysis stage, the massive amount of missing data due to sampling procedures (i.e., multistage, stratified, and unequal probability), assessment design (each student took only one test booklet or a subset of the entire test items), and non-responses from participants could negatively affect accuracy of statistical results, regardless of missing data treatment methods (imputation of missing data or deleting all missing data).

Finally, the threats to validity of the study at the data interpretation stage could stem from the variation in operationalization of the same constructs among participants and across countries. For example, for some teachers, classroom activities included working on homework; whereas for other teachers, classroom activities were restricted to only school work. Similarly, instructional time could be defined as actual teaching time in some countries and as teaching time and class management in other countries. Last but not least, the results of this study are based on the relationship between student mathematics test data and contextual data which were self-reported by students, teachers, and school principals. Self-reported data, according to Rosenberg, Greenfield, and Dimick (2006), have several potential sources of bias such as selective memory (remembering or not remembering experiences or events that occurred sometime in the past), telescoping (recalling events that occurred at one time as if they had occurred at another time), and social desirability (reporting behaviors that tend to be widely accepted by certain social groups rather than the behaviors actually exhibitted by the respondents). Thus, it is important to interpret findings of this study in light of these limitations.

## Definitions

*Developed country:* According to the World Bank's (2007) world development indicators, developed countries refer to countries with high-income economies. The use of the term, developed country, however, is not intended to imply that developed economies have reached a preferred or final stage of development (The World Bank, 2007).

*Developing country:* According to the World Bank's (2007) world development indicators, developing countries refer to countries with low-income and middle-income economies. The use of the term, developing country, however, is not intended to imply that all economies in the group are experiencing similar development (The World Bank, 2007).

*Math achievement:* For this study, math achievement is defined as the overall mathematics scores of eighth-grade students who participated in the TIMSS 2003 assessment. The overall mathematics scores can be computed by averaging students' scores on five mathematics domain contents: algebra, number, data, geometry, and measurement.

*Eighth-grade students:* In the TIMSS study, eight-grade students are defined as all students enrolled in the upper of the two adjacent grades that contained the largest proportion of 13-year old students at the time of testing (TIMSS, 2003).

# CHAPTER TWO

# LITERATURE REVIEW

## Introduction

In this chapter, the literature review is presented in six major sections: history of international mathematics achievement assessments, theoretical framework, student-related factors and student achievement, instructional practices-related factors and student achievement, teacher background-related factors and student achievement, and school background-related factors and student achievement. Finally, the chapter concludes with a summary of significant findings synthesized from this comprehensive literature review.

In broadening the scope of this literature review, the literature search was open to various student achievement outcomes such as math, science, reading, literacy, and civics. Also, empirical studies that examined student achievement outcomes at different grade levels (e.g., kindergarten to grade 12) and across countries were included in the review. It is important to note, however, that although the current literature search allowed for a broad inclusion of empirical studies related to student achievement, where possible, this synthesis of literature focused more on student mathematics achievement at middle school grades within the United States and across countries. The rationale for this selection focus was specified in Chapter One.

## History of International Mathematics Achievement Assessments

Although the beginning of internationalism in education might be traced to "ancient times" the idea of conducting an official, large-scale international achievement study did not emerge until after World War II (Encyclopedia of Educational Research [EER], 1960, p. 618). In fact, the first large-scale international achievement assessment, Pilot Twelve-Country Study, was developed in 1959 with extensive support from the United Nations Educational, Scientific and Cultural Organization (UNESCO) (EER, 1960). With the Pilot Twelve-Country Study, UNESCO aimed to promote the conviction that “educational systems cannot be transferred from one country to another, but ideas, practices, and devices developed under one set of conditions can always prove suggestive for improvement even where the conditions are somewhat different” (EER, 1960, p. 621).

The Pilot Twelve-Country Study was originally constructed in French, English and German and then translated into eight languages by individual participating countries. The test was administered in 1961 to representative samples of 13-year-old students across the 12 countries, including Belgium, England, Finland, France, Federal Republic of Germany, Israel, Poland, Scotland, Sweden, Switzerland, United States and Yugoslavia. This study assessed students’ achievement in five subject areas: mathematics, reading comprehension, geography, science and non-verbal ability. In addition, the test had two specific aims: (1) to investigate whether some indications of the intellectual functioning could be deduced from the patterns of student responses across countries; and (2) to discover the possibilities and the difficulties attending a large-scale international study (Forshay et al., 1962).

The success of the first large-scale international achievement assessment shed new light on international education. As Forshay (1962) put it, “If custom and law define what is educationally allowable within a nation, the educational systems beyond one’s national boundaries suggest what is educationally possible” (p. 7). Within approximately

50 years of development, 29 large-scale international achievement assessments were conducted, covering a vast array of subject areas including Math, Science, Reading, English, Literature Education, English as a Foreign Language, French as a Foreign Language Education, Writing, and Civic Education (International Association for the Evaluation of Educational Achievement [IEA], 2007; International Assessment of Education Progress [IAEP]; Organization for Economic Co-operation and Development [OECD], 2007). The target populations of these assessments were also expanded to students of fourth, eighth, and twelfth grades in all countries.

The popularity of international achievement assessments was also reflected through the increasing number of participating countries over the years. By decade, the largest number of countries participating in an international achievement assessment increased from 12 in the 1960s, to 19 in the 1970s, to 24 in the 1980s, to 46 in 1990s and 60 in 2000s (IEA, 2007). These powerful numerical indicators suggest that international achievement assessments have quickly gained special attention in education across countries.

## Importance of International Mathematics Achievement Assessments

A review of the history of international achievement assessments yielded an interesting finding. Of the 29 international achievement assessments conducted by IEA, 13 were mathematics assessments (IEA, 2007). In fact, since 1995, the Trends in International Mathematics and Science Study (TIMSS) has been implemented regularly, on a four-year cycle basis (TIMSS, 2007). It is worth noting that the number of countries participating in the TIMSS has also grown significantly over time. In the most recent administration of the TIMSS in 2007, more than 60 countries participated in the study, making it the largest and most ambitious international achievement study in the history of international achievement assessments.

Why have international mathematics achievement assessment attracted more attention from countries around the world? The chief reason behind the importance of international mathematics achievement assessment is not new and has been discussed for decades. Much research has linked student mathematics achievement with the future economic power as well as security of a country (Akiba, LeTendre, & Scribner, 2007; Baker & LeTendre, 2005; Carter & O'Neill, 1995; Heyneman & Loxley, 1982, 1983; Wobmann, 2003). For this reason, the differences in student mathematics achievement across countries were often interpreted as a national issue rather than a mere comparison of student achievement. For example, in the United States, it was not uncommon for national leaders to address the issue of students' poor performance in international mathematics assessments in the national agendas (see the nation's response to Sputnik crisis in the 1960s [EER, 1960], to A Nation at Risk in 1983 [National Commission on Educational Excellence, 1983] and then to Goals 2000 in 1994 [Goals 2000, 1994]). Recently, President George W. Bush, after taking office in 2001, stated: "Quality education is a cornerstone of America's future and my Administration, and the knowledge-based workplace of the 21st century requires that our students excel at the highest levels in math and science." (Bush, 2001, p.1,). As a result of such national addresses, a series of educational policies were issued in order to improve students' performance in mathematics. With the current Bush's administration, the national act of "No Child Left Behind" was implemented as a primary solution for the improvement of educational quality in the United States.

Across countries in the world, the concerns about what students know and can do in math as well as what can be done to improve student math ability has also been addressed at the national level (Beaton, 1998). As highlighted in the TIMSS 1999 Benchmarking report, the differences in students' performance in international mathematics achievements were taken seriously by many countries: "TIMSS results, which were first reported in 1996, have stirred debate, spurred reform efforts, and provided important information to educators and decision makers around the world" (TIMSS 1999 Benchmarking Introduction, 1999, p.16). Results from international mathematics achievement assessments were used for many purposes, including making changes in educational policies, setting performance standards for students, comparing with and validating national mathematics assessments, and conducting various educational research studies (Baker & LeTendre, 2005; O'Leary, 2002; Rodriguez, 2004; TIMSS, 2003).

Given the importance and profound impact of student mathematics achievement on national economic growth and security (Akiba, LeTendre, & Scribner, 2007; Baker & LeTendre, 2005; Carter & O'Neill, 1995; Heyneman & Loxley, 1982, 1983; Wobmann, 2003), and the foreseen rapid changes within and across countries in the 21$^{st}$ century, it is important that educational researchers across the world, collaboratively and separately, continuously conduct empirical research to identify factors associated with student mathematics achievement so as to maximize student learning in mathematics.

## Theoretical Framework

In an attempt to explain why students succeed or fail in their learning at school, John B. Carroll developed A Model of School Learning in 1963 (Carroll, 1963). This model proposes that the student will succeed in learning a given task to the extent that he/she actually spends the amount of time he/she needs to learn the task, with time defined as the time during which the student actively engaged in his/her learning (Carroll, 1963). According to Carroll (1963), there are five categories of variables which are associated with student's success in learning: (1) Aptitude – the amount of time needed to learn the task under optimal instructional conditions, (2) Ability to understand instruction, (3) Perseverance – the amount of time the learner is willing to engage actively in learning, (4) Opportunity to learn – time allowed for learning, and (5) Quality of instruction – the extent to which instruction is presented so that no additional time is required for mastery beyond that required in regard to aptitude (Carroll, 1963).

The five categories of variables, which can be expressed in terms of time, can be worked into a formula with degree of learning as a function of the ratio of the amount of time a student actually spends on the learning task and the total amount of time the student needs to learn the task. Thus:

$$Degree\,of\ learning = f\left(\frac{Time\ actually\ spent}{Time\ needed}\right)$$

The numerator of this fraction is equal to the smallest of the three quantities: (1) opportunity to learn, (2) perseverance, and (3) aptitude after adjustment for quality of instruction and ability to understand instruction. The last quantity, aptitude, is also the denominator of the fraction (Carroll, 1963).

Inferring from the Model of School Learning, the first three categories of variables (i.e., aptitude, ability to understand instruction, and perseverance) are related to the students; whereas the last two categories of variables are concerned with external conditions (i.e., opportunity to learn and quality of instruction). It is worthy of note, however, that of these categories of variables, opportunity to learn, quality of instruction, and perseverance are more amenable to intervention and manipulation than aptitude and ability to understand instructions which tend to be relatively resistant to change (Carroll, 1963).

## Student-related Factors and Student Achievement

### *Students' Gender*

Evidence accumulated through multiple research studies suggest that universally, a gender gap exists in math achievement (Beaton et al., 1996; Mullis et al., 2000; Peterson & Fennema,1985; Rodriguez, 2004). However, the size and direction of achievement gap varies across samples of students and tests. For examples, Bielinski and Davison (2001) found that the gender gap, albeit small, favors females in elementary school, and males in high school, and neither group in middle school. In contrast, Fennema et al. (1998) observed that the gender gap in math achievement increases during middle schools and becomes profound at the higher educational level. Generally, research findings in this area support the view that boys tend to perform better than girls on mathematics tasks such as problems that include spatial representation, measurement, proportions as well as complex problems; whereas girls tend to score higher on computations, simple problems and graph reading (Beaton et al., 1996). Similarly, using the Scholastic Aptitude Test (SAT) and classifying the math items into six levels of cognitive complexity (i.e., a zero was assigned to items measuring recall of factual knowledge, and a five to items requiring application of higher mental processes), Harris and Carlton (1993) reported that females outperformed males on the three lowest levels, whereas males outperformed females on the two highest levels, after the total test scores were controlled. When these SAT math items were grouped into two categories, applied or real world items and abstract or text-book items, the researchers found that females outscored males in abstract items; whereas males outscored females in applied items (Harris & Carlton, 1993).

Much effort has also been devoted to investigating reasons that are associated with differences in mathematics achievement between boys and girls in schools. For example, Davis and Carr (2001) suggest that the differences in use of strategies to tackle math problems in early elementary school age girls and boys are related to their achievement gap. Their study showed that boys are more likely to retrieve information from memory and use covert cognitive strategies, such as decomposition; whereas girls are more likely to use overt strategies such as counting on fingers or manipulative strategies to solve mathematics problems. Test item format is another factor that has often been linked with the gender gap in math achievement. Bolger and Kellaghan (1990), for example, have shown that boys perform better than girls in multiple-choice items and girls perform relatively better than boys in open-ended items. Findings from a more recent study conducted by Wester and Henriksson (2000), however, did not support this conclusion. In fact, Wester and Henriksson (2000) found that there was no significant change in gender differences when the item format was altered. Females seemed to perform slightly better than males when using multiple-choice items. Finally, using three nationally representative achievement databases, Bielinski and Davison (2001) examined test item difficulty as a plausible reason of gender gap in math achievement. Evidence from this study suggested an association between item difficulty and sex differences.

That is, easy test items tended to be easier for females than males, and hard test items tended to be harder for females than males. Therefore, if a math test consists of more easy items than hard items then females will outperform males in such a test, and vice versa.

### *Students' Self-confidence*

Many research studies have investigated the relationship between student self-confidence in learning math and student math achievement. For example, evidence from the study of House (2006) suggested that higher self-confidence in learning math was significantly associated with higher math achievement in adolescent students. Similarly, based on a study of middle school students in Germany, Koller, Baumert and Schnabel (2001) concluded that students with higher initial levels of interest in learning math were more likely to enroll in higher math courses. Likewise, there was also evidence that self-efficacy in learning math was significantly related to math achievement in middle school students (Pajeres & Graham, 1999). It is worth noting, however, that such a clear and positive association between student self-confidence in learning math and math achievement tended to be observed more frequently within countries. At the between-country level, the relationship between student self-confidence in learning math and math achievement appeared to be more complex. For instance, whereas self-confidence in learning math was found strongly and positively related to math achievement for students in Norway and Canada, it was not the case for students in the United States (Ercikan, McCreith & Lapointe, 2005). A similar pattern of results was also reported in the study of Mullis, Martin, Gonzalez and Chrostowski (2004) where the four countries with the lowest percentages of students in the high self-confidence category (i.e., Chinese Taipei, Hong Kong SAR, Japan, and Korea) all had high average math achievement. Likewise, in examining the relationship between math achievement and students' self-perceived competence in learning math across 38 countries that participated in the TIMSS 1995, Shen and Pedulla (2000) and Shen (2002) have shown that a negative relationship between self-perceived competence in learning math and math achievement was present between countries.

In an attempt to explain such interesting patterns of relationship, Mullis, Martin, Gonzalez and Chrostowski (2004) suggested that in Asian Pacific countries, students may share cultural traditions that encourage modest self-confidence and thus, they tended to rate themselves low in self-confidence in learning math but performed high in math assessments. Congruent with this explanation were findings from a study of Leung (2002) where the researcher observed that Japanese students tended to report more often that they were not doing well in math even though they scored high on mathematics tests. The researcher attributed such interesting patterns of students' responses regarding their level of self-confidence in learning math to the unique culture in this region of the world where the expectations for student achievement in math tend to be high (Leung, 2002).

### *Students' Valuing of Learning*

Students' valuing of learning, as defined by Ma and Kishor (1997), refers to students' affective responses to the easy or difficult as well as the importance or unimportance of a certain school subject. In existing literature, students' valuing of learning is also referred to as students' attitudes, or beliefs, or perceptions towards learning. Thus, statements such as "I enjoy learning math" or "I think learning math will help me in my daily life" can be defined as students' valuing of math or students' attitudes towards learning math.

Students' valuing of learning has often been viewed as an important determinant of student achievement. As indicated by Ma and Kishor (1997):

> Teachers and other mathematics educators generally believe that children learn more effectively when they are interested in what they learn and that they will achieve better in mathematics if they like mathematics. Therefore, continual attention should be directed towards creating, developing, maintaining and reinforcing positive attitudes (p. 27)

Empirical evidence, however, has shown little consensus concerning the relationship between students' attitudes toward learning and student achievement. Abu-Hilal (2000), for example, asserted that students' perceptions regarding the importance of mathematics exerted a significant effect on math achievement. Similarly, findings from the study of Beaton et al. (1996) revealed that eighth grade students with more positive attitudes had higher average mathematics achievement. In a meta-analysis study, Ma and Kishor (1997) examined 113 studies that investigated the effects of students' attitudes on math achievement for the period from 1966 to 1993 and found that approximately 90% of the studies showed positive relationship between attitudes and achievement. The overall weighted mean effect size obtained from this meta-analysis study was 0.12, with a 95% confidence interval from 0.12 to 0.13, suggesting a positive, albeit not strong, relationship between attitudes and math achievement. More evidence supporting significant relationships between the value students attach to math and their achievement in math can also be found in Marsh, Hau, and Kong (2002), Rao, Moely, and Sachs (2000) and Singh, Granville, and Dika (2002).

Opposing this view, Shen (2002) conducted a study on the relationship between eighth grade students' achievement and their self-perception of learning math across 38 countries that participated in the TIMSS 1995 and concluded that empirical evidence is insufficient to support the claim that attitudes and achievement are strongly related. Although the researcher did find a positive relationship between math achievement and three measures of self-perception (i.e., how much students like the subject, their self-perceived competence in the subject, and their perceived easiness of the subject) for within-country data, the between-country analysis yielded opposite findings. That is, there was a negative relationship between self-perceptions and achievement. The correlation between math achievement and how much students like mathematics cross-nationally is -.68. The two countries with the highest scores for liking math (i.e., Morocco and South Africa) were also the countries that had the poorest performance in math. When correlating students' perceived easiness of math and math achievement across countries, a stronger negative correlation (-.72) was observed, indicating that in poor performing countries students were likely to think of math as being easy whereas in high performing countries, students were likely to think of math as being difficult. Explaining this negative pattern of relationship between math achievement and students' self-perception, Shen (2002) suggested that this pattern might reflect low academic standards and expectations in low performing countries and high academic standards and expectations in high performing countries.

Evidence from cross-national studies of Papanastasiou (2000, 2002) also did not support the contention that positive attitude towards learning is associated with greater student achievement. Of the three countries included in Papanastasiou's (2000) study (i.e., Cyprus, the United States, and Japan) Cyprus had the highest proportion of student who reported positive or strongly positive attitudes toward learning mathematics (79% as compared to 70% for the U.S. and 51% for Japan). However, Cyprus students' average mathematics score was the lowest among the three countries (474 as compared to 500 for the U.S. and 605 for Japan). Later, in 2002, using the TIMSS 1999 data, Papanastasiou replicated this study on the samples of students from Cyprus, Hong Kong, and the United Stated. Interestingly, the new study yielded similar results. That is, having the largest proportion of students reporting positive or strongly positive attitude towards math did not make Cypress the country with the highest average math score.

In a recent international study, House (2006) looked at the relationship between students' attitude towards math and math achievement in the TIMSS 1995. Fourth grade students from the United States and Japan who participated in the TIMSS 1995 were included in the study. Three attitude-related variables (i.e., I enjoy learning math, math is boring, and math is an easy subject) were simultaneously entered in a multiple regression model to predict student math achievement in each country. Results from this study indicated that "I enjoy learning math" had a statistically significant positive relationship with math achievement in Japan but not in the United States. Put differently, students in Japan who earned high math scores also tended to indicate that they enjoyed learning math; whereas the same relationship was not significant when tested with the sample of United States students. The researcher, however, noted some similarity between the two countries. That is, in both Japan and the United States, "math is boring" was significantly negatively related to student math scores. Specifically, students who expressed the belief that math was boring also tended to achieve low math test scores. Although "math is an easy subject" appeared to have a negative relationship with math achievement in Japan and a positive relationship with math achievement in the United States, these relationships were not statistically significant in either of the countries.

In another international study, using the sample of Swedish eighth grade students ($n$ = 343) participating in the TIMSS 2003, Eklof (2007) examined the relationship of math achievement with students' value of math, math self-concept and test-taking motivation. In this study, value of math was a composite variable which was computed as a mean score of six indicators and ranged from 1 (strongly agree) to 4 (strongly disagree). Results of the multiple linear regression analysis suggested that altogether the three predictors explained about 31% of the variation in the student math scores. However, when the effects of math self-concept and test-taking motivation were partialled out, the relationship between value of math and math achievement for the sample was weak and negative in direction (Eklof, 2007).

### *Students' Family Background*

Following the Coleman report (1966) which suggested strong evidence that home background-related factors had significant effects on student learning, extensive research has been carried out, both in the United States and in other countries, to validate Coleman's findings (Baker, Goesling, & LeTendre, 2002; Comber & Keeves, 1973; Coleman, 1975; Heyneman & Loxley, 1982; Fuller, 1987; Suter, 2000). Home background here refers to a vast array of factors including, but not limited to, parental education level, family socioeconomic status, family size, and home resources. Unfortunately, results from these studies shared little agreement.

Three studies that provided conflicting results with Coleman (1966) are Comber and Keeves (1973), Heyneman and Loxley (1982), and Fuller (1987). For the first two studies, the researchers used the same data source [i.e., IEA's (1971) First International Science Study (FISS)] for 18 countries to examine the relationship between home- and school-related factors and student achievement in science. It is worthy of note that the 18 countries included in the study consisted of both developed and developing countries. In the study of Comber and Keeves (1973), the researchers employed a three-step data reduction process to select variables for the analysis. First, only background variables that had correlation coefficients with achievement larger than twice their standard error were considered. Second, the effects of background variables were partialed out before other teacher- and school-related variables were entered in a regression model. The resulting standardized regression coefficients were then calculated for each background variable. Finally, the standardized regression coefficient was averaged across 18 countries and those that exceeded .05 were included in subsequent analyses. Results from this study showed that across countries, teacher- and school-related variables exerted stronger positive effects on student achievement than family background variables.

Arguing that Comber and Keeves' (1973) method for variable reduction was essentially flawed because it assumed that a background variable had to be a strong predictor of science achievement in both developed and developing countries to be included in the final analysis, Heyneman and Loxley (1992) applied a new procedure of submitting each potential variable to the same test of importance, but in each country separately. As a result, the list of variables to be included in Heyneman and Loxley's (1992) step-wise regression analysis varied from one country to another as opposed to the same list of selected variable applied uniformly to all countries in the Comber and Keeves' (1973) study. As suggested by this study, the poorer the country in economic status, the more impact teacher and school-related variables seemed to have on student achievement.

For the third study, Fuller (1987) used data from only developing countries to investigate the association between family background and student achievement. The statistical method employed in Fuller's (1987) study was multiple regression analysis. Results from this study revealed that in developing countries, the effects of family background on student achievement were non-significant relative to the effects of school. In fact, Fuller (1987) found that in India the effects of school explained up to 90% of the variance in student achievement.

Contradicting with the conclusion made in the studies of Comber and Keeves (1973), Heyneman and Loxley (1982), and Fuller (1987), were findings from a more recent study of Baker, Goesling, and LeTendre (2002). Evidence from this study suggested that the relationship between family SES, a composite variable of mother's and father's education level and number of books in the home and student achievement were similar across countries, regardless of national income. For this study, the results were obtained from the analysis of the TIMSS 1995 data for both eighth grade math and science using hierarchical linear modeling (HLM) technique to control for the nesting structure of the data. A total of 36 countries with both low and high economic development status were included in the study.

Most recently, using the TIMSS 2003 eighth grade mathematics data, Mullis, Martin, Gonzalez and Chrostowski (2004) studied the association between student home resources and their math achievement. Once again, their findings did not agree with those of Baker et al. (2002). In particular, Mullis et al. (2004) highlighted that in many countries, students from homes with a range of study aids such as computer, calculator, desk, and dictionary had higher achievement in math than their peers who did not have access to such resources at home.

### *Students' Time on Homework*

Time student spent on homework is a key variable in Caroll's model for school learning (Caroll, 1963) and subsequent homework studies. There is ample evidence that time on homework is positively related to students' academic performance (Cooper, 1989a; Cooper, Lindsay, Nye, & Greathouse, 1998; Cooper & Valentine, 2001; Keith & Cool, 1992; OECD, 2001; Peterson & Fennema, 1985; Singh, Granville, & Dika, 2002). Of these studies, the work of Cooper (1989a) has been widely cited in existing literature. Cooper (1989a) reviewed approximately 120 studies on the effects of homework conducted between 1962 and 1987 which tended to fall in one of the two types of research designs: experimental and quasi-experimental.

In the 50 studies that specifically examined the relationships between time on homework and academic achievement, Cooper (1989a) noted that time spent on homework was operationalized as time spent on homework per week. Regarding achievement measure, the majority of these studies used standardized tests (33 studies), some used class grades (7 studies) and some used other outcome measures such as motivation to learn (10 studies). Statistical methods employed in these studies included structural equation modeling, path analysis, and repeated measures ANOVA. As a result of this comprehensive review, Cooper (1989a) concluded that most research showed a positive relationship between the amount of time spent on homework and academic achievement. However, the effects of time on homework were larger for middle and high school students and near zero for elementary school students (Cooper, 1989a).

In an attempt to further elucidate the positive relationship between time on homework and achievement, Singh, Granville, and Dika (2002) examined the effects of motivation, attitude, and academic time on math achievement by building a latent variable structural equation model via a confirmatory factor analysis approach. In this study, academic time was measured by time students spent on math homework and time students spent watching television on the weekdays. The model was fit to the sample data of 24,599 eighth graders in the United States who participated in the National Education Longitudinal Study 1988 (NELS:88). Listwise deletion method was adopted to handle missing data on variables of interest and mathematics test scores. The resultant sample of 3,227 students was used for subsequent analyses. Findings from this study supported the positive effects of the three factors: motivation, attitude, and academic time on math achievement. Specifically, in examining the measurement part of the model, the researchers found that time on math homework was the better indicator of academic time than time watching TV on weekdays. In terms of structural relations in the model, of all the latent variables (motivation, attitude, and academic time), academic time had the strongest direct effect on math achievement ($\beta = .50$ as compared to .23 for attitude, and .16 for motivation). Altogether, this model accounted for 46% of the variance in math achievement (Singh, Granville, & Dika, 2002).

Despite the history of homework research that supports the positive relationships between time on homework and math achievement, several researchers recently argued that such findings are questionable. In reviewing previous studies, Trautwein and Koller (2003) found two common major pitfalls: (a) confounding operationalization of time on homework and (b) problematic handling of hierarchically ordered data. According to the researchers, time on homework was defined very differently across studies, from time spent on homework per week which could mean time students spent on homework in all subjects or in a specific subject to after school-related activities (Trautwein & Koller, 2003). What's more, in some studies, time on homework was an aggregated variable which consisted of homework frequency (i.e., frequency of homework assigned by the teacher, a class-level variable) and homework length (i.e., the time typically spent on homework per day, a student-level variable). Even when disaggregated, time on homework as reported by individual students might have different meanings ( e.g., time to complete the whole homework vs. part of the homework) (Trautwein & Koller, 2003).

On the statistical analytic methods of previous studies, Trautwein and Koller (2003) observed that two possible effects of time on homework (i.e., student-level effect and class-level effect) were often mixed up. For example, in examining homework effects, Trautwein and Koller's (2003) found that, in previous studies, "no homework is ever required" (teacher effect) and "I have homework, but I don't do it" (student effect) were collapsed into a single response category (p. 122). As a result, the complex measure of time on homework was often treated exclusively at only one level, either student-level or class/teacher-level (Trautwein & Koller, 2003). Such a conceptual model is problematic because it neglects the non-independence of individual student data and would likely lead to biased estimates of several statistical parameters such as fixed and random effects, as well as an inflation of type I error rate (Bryk & Raudenbush, 2002).

According to Trautwein and Koller (2003), the solution for overcoming the identified shortcomings of statistical analyses in previous studies was to differentiate between the student-level effects and the class/teacher effects and simultaneously conceptualize time on homework at both of the levels as is possible in multilevel modeling (Bryk & Raudenbush, 2002).

Recently, Trautwein (2007) re-analyzed the Programme for International Student Assessment (PISA) 2000 data used in the Organization for Economic Co-operation and Development (OECD) (2001) study that supported the view that longer homework time is associated with higher achievement in math.  Homework time in PISA 2000 had four response categories: no time, less than one hour a week, between 1 and 3 hours a week, and 3 hours or more a week. Using multilevel modeling to account for student-level and school-level time on homework effects, Trautwein (2007) found that the relationship between homework time and achievement was only moderate at the school level and was negative at the student level. In other words, students who spent more time on math homework than their peers scored lower on the math assessment, whereas a high average homework time at the school level was positively related to achievement.

Congruent with Trautwein's (2007) results were findings from Rodriguez's (2004) study which showed that the amount of time students spent on homework was negatively related to math performance, after holding other variables constant. For this study, a two-level Hierarchical Linear Modeling (HLM) model was employed to analyze the sample data which included 328 teachers and 6,963 eighth grade students from the United States who participated in the TIMSS 1999 study. Time on homework was dummy coded into two variables to capture three levels of student effort: (a) no homework, (b) more than 0 and up to 1 hour of homework, and (c) more than 1 hour of homework (Rodriguez, 2004). Findings from this study suggested that students who did no homework each day performed slightly higher on average than those students who spent more than 1 hour a day on math homework. Similarly, comparing to their peers who spent more than 1 hour on math homework each day, those students who spent about 1 hour doing homework tended to have higher math scores. In explaining this inverse pattern of relationship between time on math homework and math achievement, Rodriguez (2004) argued that the students who spent more than 1 hour each day studying math were likely poor math achievers and thus, they needed to study more to catch up with their peers.

### *Academic Tutoring*

Academic tutoring has become more common for students around the world. The reason for student engagement in this activity is primarily academic improvement. However, there seems to be insufficient empirical evidence to support the belief that academic tutoring consistently and positively increases student learning. In fact, many researchers are debating whether academic tutoring is beneficial for students of various ability levels. Because academic tutoring takes many shapes and forms, the task to identify which academic tutoring programs are related to increasing student achievement is relatively challenging.

In the United States, the past recent decades have seen states, local districts, and schools focusing on making more time for teaching and learning (Yair, 2000). This was in response to the report of the National Commission on Excellence in Education (NCEE), A Nation at Risk, that American students were repeatedly under performing in

Math and Science compared to students of other countries (National Commission on Educational Excellence [NCEE], 1983). Academic tutoring as a result of "innovative manipulations of time" turned out to be one of the plausible solutions that schools could implement in the hope of improving their student achievement (Yair, 2000, p. 485). Examples of innovative manipulations of time include using lunch breaks for tutoring programs, rescheduling school bus time, and cutting back sports and other extracurricular activities so as to give students more time to do homework in school (Yair, 2000).

Cosden, Morrison, Albanese, and Macias (2001) reviewed nine studies that examined the effects of two types of at-school tutoring programs, homework assisted and academic enrichment, on student achievement. This study found that students who participated in tutoring programs that offered academic enrichment (e.g., literacy skill building, mathematics adaptive skills training, reading with specialist, etc.) tended to have higher achievement test scores in reading, language, and math than did children in the control groups. However, for tutoring programs that offered homework support, the results were mixed. For example, students in the study of Beck (1999) reported that their participation in the tutoring program where they were provided with time, a structured setting for homework completion, and instructional support had increased their performance at school. In contrast, evidence from the study of Ross et al. (1992) indicated that extended homework time for elementary children at tutoring programs was counterproductive in terms of student performance on standardized tests. It is important to note here that all of the children participating in these tutoring programs were identified as minority students or students at-risk for school failure.

Consistent with these findings, Adler (1996) stated that minority students were often provided with additional study hours relative to White and Asian students. Schools implemented homework-at-school programs, enrichment programs, and extra academic or language classes to allow minority students to catch up more easily with their majority peers. In fact, educational and political activists had joined forces to provide more funds in terms of hours for these programs because time to learn was viewed as the most valuable asset in reducing ethnic, racial, and social inequalities in education (Adler, 1996).

A similar educational strategy (i.e., provision of additional academic support to low-achieving students and students of special needs) has also been well supported in England (Muijs & Reynolds, 2003). However, the quasi-experimental study conducted by Muijs and Reynolds (2003) did not yield favorable results. The study found that first grade and second grade students who had received at-school academic tutoring did not make more progress in mathematics than those who had not. As such, this study did not provide evidence to support academic tutoring as a way of improving the achievement of low-achieving students.

In addition to on-site academic support, there are other types of academic tutoring that take place outside of the school setting. These programs include private tutoring, training for spelling bees, and preparation for quizzes and exams, to name a few. According to Lee (2007), private tutoring is a preferable form of tutoring in many countries. The focus of private tutoring, however, differs considerably across nations. For example, in Korea, private tutoring serves primarily enrichment needs for higher achieving students; whereas private tutoring in the United States is primarily for meeting remediation needs of lower achieving students (Lee, 2007). Still, for some other countries such as Mainland China and Hong Kong, private tutoring serves mainly to supplement students with what has been missed in the regular classrooms. As Bryan et al. (2007) pointed out, in Mainland China and Hong Kong where class sizes were often large, hands-on explorations became difficult and individual care and guidance was often left to after-class hours.

With regard to the impact of these private tutoring programs on student achievement, it appears that the example of Japanese students was often used to argue for the benefits of private tutoring. As indicated in the study of Walberg and Paschal (1995), in 1993, roughly 42% of sixth graders, 53% of seventh graders, and 59% of eighth graders in Japan attended additional lessons in private supplementary schools where students had the opportunity to do exercises relevant to their schoolwork and to learn mathematical skills. As a result of these programs, concluded by Walberg and Paschal (1995), Japanese students consistently performed on the top of international achievement tests. Such a conclusion, however, should be interpreted with cautions because in countries such as Japan and Korea, the majority of the students who participated in private tutoring were of medium- to high-ability levels who perceived private tutoring as a tool that could help them exceed beyond the norm.

In another international study, Papanastasiou (2002) used the TIMSS 1995 data to investigate the effects of extra math lessons on fourth grade student achievement in Cyprus, Hong Kong and the United States. Dissimilar to findings from prior research, this study found that among all the students in the study, those who responded that they did not take extra math lessons had the highest math scores, and this pattern of relationship was consistent across the three countries. According to Papanastasiou (2002), the results were somewhat expected because extra math lessons would most likely be needed by the students who are not strong in math.

## Instructional Practices-Related Factors and Student Achievement

Around the world, classrooms are the common places where students come to learn and construct new knowledge, develop competences and prepare themselves for the future. Instructional practices refer to those activities that are designed and implemented by teachers in the classrooms in order to maximize student achievement. According to Cogan and Schmidt (1999), instructional activities are one of the most important factors related to student learning in the classrooms. The following section presents results from studies that investigated instructional practices in four areas: opportunity to learn mathematics, activities in math lessons, mathematics instructional hours, and amount of math homework and their association with student achievement in mathematics.

### *Opportunity to Learn*

Although opportunity to learn was identified as one of the five key elements in the model for school learning (Caroll, 1963), it appears that little research attention was paid to elucidate the effect of opportunity to learn on student achievement (Pianta et al., 2007). Not until the 1990s, in responding to new legislation related to Goals 2000 and the Elementary and Secondary Education Act (Muthen et al., 1995), did a group of researchers in the United States begin to focus their interest on opportunity to learn as a potential factor that could help enhance student learning and improve teaching (Muthen et al., 1995; Wiley & Yoon, 1995). Research on opportunity to learn appeared to receive more attention when results from various international achievement comparative studies indicated that students from the United States performed less well in math and science than their peers from other countries such as Japan, Korea, and Singapore (Baker, 1993; Westbury, 1992). There were increased concerns about the fairness of these international comparative studies. Many researchers argued that the fairness of these international comparisons could be compromised by differential learning opportunities across schools and nations (Wiley & Yoon, 1995). It is very important to note, however, that as time progressed, the traditional definition of opportunity to learn (i.e., time allowed for learning) (Caroll, 1963) was reconceputalized to include not only time allowed for student learning but also other educational aspects such as content or curriculum coverage, instructional activities, and instructional time (Schmidt & McKnight**,** 1995).

In 1992, using the Second International Mathematics Study (SIMS) data, Westbury conducted a study to examine the effects of curriculum implementation on student achievement. To control for potential differences in curriculum or intentional emphasis on grade 7 and 8 algebra and grade 12 calculus in Japan and in the U.S., Westbury (1992) classified sampled U.S. and Japanese classrooms into course types (i.e., remedial, typical, enriched, algebra in the US. versus math in Japan) and then matched the percentage of topic coverage reported by teachers for each math topic with the number of SIMS items in the same content areas. Next, based on the opportunity to learn (i.e., percentage of topic taught in a course) students in enriched and algebra courses in the U.S. were compared with all Japanese students. Results of this study showed that where the American curriculum was comparable to both the curriculum of the SIMS test and the curriculum of Japan, U.S. achievement was similar to that of Japan. Therefore, Westbury (1992) concluded that overall, the lower achievement of the United States was the result of curriculum that was not as well matched to the SIMS tests as was the curriculum of Japan.

In addressing the same research questions raised in Westbury's (1992) study, Baker (1993) took a different approach to re-analyze the SIMS data. According to Baker (1993), Westbury's (1992) analysis, which rested on greatly restricting the American sample to control for curriculum differences, was problematic. In other words, Westbury compared "parts" of the American sample, i.e., students in enriched and algebra who tended to be high ability students, with the "full" Japanese sample. Thus, to correct for this issue, Baker (1993) compared achievement of American and Japanese students based on taught curriculum and untaught curriculum which could be computed from SIMS' teacher data.

Not surprisingly, findings from Baker's (1993) re-analysis were quite dissimilar to those of Westbury's (1992) study. Baker (1993) showed that substantial differences existed in the effectiveness of the two systems that went beyond curriculum coverage. Specifically, on average, Japanese students learned over 60% of what they were taught in the target grade, whereas their American peers learned only 40%. Additionally, three-fourths of Japanese scores were above the median of the American distribution, and over 9% of Japanese students learned all of the tested material taught to them as compared to less than 2% of American students. The effect size of the mean differences between the two systems was 0.81, which was large. In regard to untaught material, Japanese students appeared to know more material that had not been taught than American students, with an effect size for the mean difference of the two systems of 0.33. In addition, Baker (1993) also examined the distribution of the gain scores of students in the two countries after accounting for curriculum taught and untaught during target grade. It appeared that there was a wide variation in yearly performance of American classrooms, with a few doing better than the best Japanese classrooms and some showing no or even negative gain over the year. In contrast, Japan had relative small variation in yearly performance between classrooms, with a relatively high minimum level of performance. In summary, Baker (1993) concluded that the Japanese system imparted more math knowledge to more students than the American system.

To control for cross-cultural differences, Wiley and Yoon (1995) used the data from the California Learning Assessment System (CLAS) 1993 to investigate the impact of learning opportunity on student math achievement in the United States. For this study, approximately 1,750 math teachers (1,100, 420, and 230 teachers in Grades 4, 8, and 10, respectively) and 30,250 students (17,250, 10,100, and 3,000 students in Grades 4, 8, and 10, respectively) were included. The variable opportunity to learn was created based on mathematics teachers' responses on whether they actually taught mathematics to *almost all* of the students who were given the math test. In addition, teacher-related variables such as familiarity with curriculum goals and standards, participation in professional development, and implementation of instructional practices were also used. Students' test scores on the CLAS 1993 were used as the outcome variable. Results from this study suggested that students' exposure to different math topics, and the way in which these topics were covered affected their performance on tests. However, the impact of opportunity to learn differed across grade. In Grade 4, for those teachers who were familiar with mathematics instruction assessment guides, and who participated in mathematics curriculum activities, students performed significantly better than students of the teachers who were not involved in those activities. Participation in professional conferences formed the notable exception to this pattern. A similar pattern was observed in Grade 8. Grade 10 showed the least impact despite the highest level of teachers' familiarity with math goals and standards and frequent participation in various instructional activities.

Driven by the desire to further understand how mathematics was taught and learned from a qualitative perspective, Schmidt, McKnight, Valverde, Houang, and Wiley (1997) conducted a large-scale investigation of the mathematics curricular visions and aims of almost 50 countries participating in the TIMSS 1995. For the first time in a study of this scope, all the data came from national curricular documents (i.e., official curriculum guides and textbooks). Findings from this study indicated that curricular data varied substantially across countries with differences observed in the kinds of learning opportunities provided, in the mathematical content involved, in the expectations for students, and in the organizing and sequencing of the opportunities provided (Schmidt et al., 1997).

As an example of mathematics curricular coverage, although the same major mathematics topics (i.e., algebra, number, geometry, measurement, and data) were introduced in all countries at some point within the schooling years, the focus on a particular topic as well as the duration and depth of content coverage for each topic differed from one country to another in myriad ways (Schmidt et al., 1997). At eighth grade, in China, Czech, Japan, Korea, Netherlands, Portugal, Romania, Slovak, and Spain, the largest proportion of textbooks (more than 30%) was devoted to the topic of algebra; whereas in Philippines, South Africa, and Sweden, little emphasis (less than 5%

of textbooks) was placed on this topic. Similarly, at fourth grade, whereas Scotland, Romania, Belgium, the Russian Federation, Slovenia, Mexico, and Portugal gave considerable attention to the topic of measurement (more than 30% of textbooks), Israel and the Philippines featured little or no coverage of this topic (Schmidt et al., 1997). As a result of such pervasive differences in opportunities to learn across nations, Schmidt et al. (1997) concluded that it was not surprising to observe substantial differences in international student achievement. These differences in student achievement, however, must be carefully interpreted in context (Schmidt et al., 1997).

### *Homework Assignment*

Homework assignments and their effects on student achievement is another topic that has recently drawn considerable attention from educational researchers. Because the effects of homework assignments should be conceptualized as occurring at the class/teacher level (Trautwein, 2007), it is not surprising to observe that most recent studies on homework assignments (De Jong et al., 2000; Rodriguez, 2004; Trautwein et al., 2002; Trautwein & Koller, 2003; Trautwein, 2007) have adopted multilevel modeling methods for analyzing hierarchically structured data. In general, empirical evidence from these studies suggested that the frequency and amount of homework assignments were positively related to student achievement. However, the effect size of homework assignments varied across grades.

In the study conducted by De Jong et al. (2000), the researchers examined the effect of homework assignments (i.e., amount of homework assigned or the number of homework tasks and frequency of homework assignments) on class achievement in math. A sample of 28 schools, 56 classes and 1,394 middle school students from the

Netherlands was used in this study. The dependent variable of the study was student math scores from the national standardized test. The independent variables, amount of homework and frequency of homework, came from the teachers' log book during the school year. Data were also collected from students regarding homework time, homework problems, homework study tactics and the role of parents. In addition, classroom observation data were gathered regarding homework assignment and homework discussion. In terms of data analysis, the relationships between homework assignment aspects and achievement were analyzed using correlations, partial correlation, and multilevel modeling technique. Results of this study suggested that the amount of homework (i.e., the number of tasks assigned to the class in the respective school year) was significantly related to class achievement (this variable explained 2.4% of the variance in class achievement) whereas the frequency of homework had no effect on math achievement at the class level and time on homework had no effect on student math achievement. According to the researchers, a possible explanation for the non-significant effect of frequency of homework was probably the restricted variance in homework frequency. That is, a large majority of teachers in the Netherlands assigned homework to students relatively frequently.

Another recent study using multilevel analyses to investigate the effects of homework assignment on class achievement was Trautwein et al. (2002). Using a subsample of a large-scale national database of adolescents in Germany, Trautwein et al. (2002) looked at the relationship of homework variables on math achievement in school. Specifically, in this study, repeated measurement data were collected from 1,976 seventh grade students from 125 classes at two time points (i.e., at beginning of the school year and at the end of the school year). Regarding key variables of the study, math achievement was measured by students' standardized test scores for the 1991-1992 school year. Homework variables included frequency of homework assignment (an aggregated variable of homework frequency and frequency of teachers' monitoring of homework completion at the class level) and time typically spent on homework per day, a student-level variable. Results of the multilevel modeling analysis suggested that the frequency of homework assignment was positively related to math achievement (the explained variance at the class level, after controlling for other variables in the model, was 8%); whereas time on homework had no significant effect on math achievement. Interestingly, Trautwein et al. (2002) also found that, in this study, the positive effect of frequency of homework assignment did not depend on whether homework assignment was short or long.

In the United States, Cooper (1989a) conducted a comprehensive review of studies on the effect of homework on academic achievement. To give a focus for his study, Cooper defined homework as "tasks assigned to students by school teachers that are meant to be carried out during non-school hours" (p. 7). This definition excludes (a) in-school guided study, (b) home study courses, and (c) extracurricular activities (Cooper, 1989a). In the area of homework assignment, Cooper (1989a) reviewed 20 experimental studies which compared the achievement of students given homework assignments with students given no homework or any other treatment to compensate for their lack of homework. These studies represented all levels of education, from elementary schools to middle and high schools.

As a result of this study, Cooper (1989a) concluded that homework assignment had positive effect on student academic achievement but the size of effect varied across grades (Grades 4–6: $d = 0.15$; Grades 7–9: $d = 0.31$; Grades 10–12: $d = 0.64$). This was based on evidence that 14 of the 20 studies produced effects that support homework assignments. Interestingly, the researcher noted that there was no clear pattern indicating that homework was more effective in some subjects than in others. For middle school students, homework assignment was significantly related to student achievement. The amount of homework assignment was optimal when it required between 1 hour and 2 hours per night to complete. For high school students, more homework assignments were associated with better student achievement. As an example, an average high school student in a class with homework assignments tended to outperformed 69% of the students in a no homework class. At the elementary level, however, the effect of homework assignments, regardless of amount, tended to be negligible (Cooper, 1989a).

### *Classroom Activities*

Prior research presents a fairly positive view toward the relationship between students' experiences in academic classrooms and their achievement (Yair, 2000). Yet, debates continue among researchers regarding types of activities and the extent to which those activities impact student achievement (Staub & Stern, 2002). Believing in how much students learn is determined by the time they actually spent on-task (Caroll, 1963), Yair (2000) studied student engagement/disengagement with instruction in various academic classrooms in the United States and found that activities such as laboratory work, small group discussions, and presentations were highly engaged by students. Activities that were teacher-directed or teacher lectures attracted the lowest rate of engagement from students. Yair (2000), therefore, concluded that more individualized activities or active instructional practices would likely result in higher student achievement. Yair's (2000) study, however, is challenged in that students may learn more not through an activity that they engage in at a higher rate but from lower engagement in an activity that is more conducive to learning. For example, a student may learn more from lower engagement in a lecture as opposed to higher engagement in group work. In addition, it is possible that misconception exists in small group discussion and, as a consequence, higher engagement in this activity is not conducive to knowledge acquisition (Yair, 2000).

For some classrooms, Cooper (1989a) observed that the type of activities that teachers preferred to do in class was to review, discuss or even allow students to do some homework-like assignment. In a meta-analysis study of homework, Cooper (1989a) reported a set of studies that compared the effect of homework with that of in-class supervised study on student academic achievement. In these studies, students not receiving assignments to complete at home were asked to complete some assignments in class. These activities typically were assigned at the end of each unit or lesson. Results of these studies suggested that the effect of homework was about half of that of in-class homework-like assignments (Cooper, 1989a). Interesting to note, however that, after controlling for grade, the effect of homework assignment tended to be significantly larger than that of in-class homework-like assignment. Such a pattern of findings was observed in studies where junior high and high school students were sampled but not in studies where elementary students were sampled (Cooper, 1989a).

Across countries, it was found that notable differences exist in the types of activities exposed to students in the classrooms. In an observational study of activities in elementary school mathematics classrooms in the United States and Japan, Stigler et al. (1987) revealed that teachers in Japanese classrooms spent significantly more class time asking academic questions of the entire group whereas teachers in the United States asked significantly more questions of individual students. Later, replicating the study in middle school mathematics classrooms, Stigler et al. (2000) found that students in Japan spent more class time on activities designed for inventing and proving and less time on practicing routine procedures than did students in the United States. Similarly, a qualitative study of Bryan et al. (2007) found that there were significant differences in activities in mathematics lessons between countries. Whereas Australian and American teachers tended to use hands-on manipulative activities frequently in math lessons, Mainland Chinese and Hong Kong teachers tended to engage students more in teacher-led whole class activities or verbal activities where students have opportunities to discuss, question, and answer. Also, there appear to be significantly more group activities and in-class student collaboration in the United States than in Australia, Mainland China, and Hong Kong (Bryan et al., 2007).

Because Japanese students consistently performed better than American students in various international achievement tests some researchers tended to attribute Japanese students' academic success to the types of activities they experienced in classrooms. Hiebert and Stigler (2000), for instance, recommended that more effective instructional strategies should be used in academic classrooms based on Japanese approaches. Bryan et al. (2007), however, argued that the types of activities implemented in Eastern schools were not necessarily more effective; rather they were strongly influenced by cultural factors. As indicated in their study, even though teachers from Hong Kong and Mainland China recognized the benefits of more individualized and hand-on manipulative activities they simply could not use these activities due to large class size and pressures to cover a heavy load of subject materials in the time assigned (Bryan et al., 2007). Stipek, Givvin, Salmon, and MacGyvers (2001), on the other hand, asserted that the type of activities used in the classroom was strongly influenced by teacher's beliefs. As evident in the one of the studies that these researchers examined, teachers who believed that children learn mathematics by constructing their understanding in the process of solving problems tended to give students more word problems in instruction and spend more time developing children's counting strategies before teaching number facts (Stipek, Givvin, Salmon, & MacGyvers, 2001).

### *Instructional Time*

Caroll's (1963) model for school learning suggested that the amount of time devoted to learning is an important determinant of how much is learned. It would be reasonable, therefore, to expect that the more instructional time is provided to students, the greater achievement is likely to result. However, existing literature suggests that this is not always the case. Cooper (1989b) attributed the unclear relationship between instructional time and student achievement to the various definitions of instructional time used in existing literature. For example, in some studies, instructional time was defined as scheduled or allocated time that was set aside by law, school, and/or teacher for a particular activity to take place; whereas in other studies, instructional time was operationalized as actual amount of time spent on academic material within the allocated time. Still in other studies, instructional time was measured as engaged time or time-on-task that excluded time for classroom management and interruption (Cooper, 1989b). For Schmidt et al. (1997), instructional time should be interpreted as time in which specific educational opportunities are made available to students within any school year. Yet, according to Baker, Fabrega, Galindo, and Mishook (2004), instructional time could be measured in several ways, from the number of days in the school year to the number of hours spent on a single subject.

Results from Evertson's (1980) study indicated that a larger amount of time devoted on-task was associated with greater student achievement. Specifically, the study found that low-achieving junior high school students tended to engage about 40% of the time; whereas high-achieving students appeared to engage about 85% (Evertson, 1980). These findings were supported by the research of Fredrick and Walberg (1980) where the researchers reviewed nine studies that examined the relationship of instructional time in terms of time-on-task and student achievement and found that all nine showed a positive relationship (Fredrick & Walberg, 1980). In the studies reviewed, the correlations ranged from .15 to .53. Even after controlling for other variables such as I.Q, ability and readiness, the correlations were still positive, ranging from somewhat weak to moderately strong ($r$ = .09 to .44). The researchers, therefore, proposed that school days or year should be lengthened in order to increase student achievement (Fredrick & Walberg, 1980).

Empirical evidence supporting the positive effects of instructional time on student achievement can also be found in one of the most influential reports in the history of United States educational reform, A Nation at Risk (NCEE, 1983). By correlating the number of school days with academic achievement, the report suggested that Asian students performed better than American students in math and science because they had more time to study. On average, Asian students studied up to 240 days a year compared with about 180 days for American students. Therefore, in order for American students to be competitive globally, NCEE recommended that United States schools should increase an extra school hour per day and up to 40 extra days per school year (NCEE, 1983).

With international educational data becoming more accessible, many researchers recently had the opportunity to examine the effects of instructional time on student achievement in a much more diverse setting. Using three major international databases: Programme for International Student Assessment (PISA) 2000, Trends in International Math and Science Survey (TIMSS) 1999, and International Study of Civic Education (CIVICS) 1999, Baker, Fabrega, Galindo, and Mishook (2004) investigated the relationship between instructional time and student learning across countries. PISA 2000 tested mathematics, science and reading skills of 15-year-old students from 32 countries; TIMSS 1999 tested mathematics and science of eighth-grade students from 38 countries, and CIVICS assessed eighth-grade students' knowledge in civics from 28 countries. In this study, test scores were the outcome variable and instructional time during the academic year, in terms of hours, dedicated to formal educational activities was the independent variable.

Results from this study indicated that there was no statistically significant relationship between instructional time and achievement scores at the cross-national level. Also, this pattern of relationship was observed consistently across the three databases and subjects tested. For example, in the TIMSS data, students attending math class for 5 hours or more during the week score 481 on achievement tests, while students who receive less than 2 hours of math per week score on average of 485. About 90% of the students who received between 2 and 5 hours of math class got an average 491 points on the math achievement test.

Interestingly, the evidence that more hours of math class did not result in better achievement scores was also observed within nations. The average statistically significant correlation for the positive effect is 0.09, meaning that the relationship between total instructional time and math achievement within a country accounted for only 0.8% of the variance in achievement scores. There were several cases where a negative effect was observed. In these cases, however, the average magnitude of the effect was also small, about 1.4%. Hungary and Japan were the only two countries where the magnitude of the effect was somewhat noticeable. For tenth grade, Hungarian students who received more than 912 hours of instruction per year tended to score 55 points higher than their Hungarian peers who received less than 810 hours of instruction. Similarly, Japanese students who received more than 1,112 hours of instruction per year scored 25 points higher than their Japanese counterparts who received less than 935 hours of instruction (Baker et al., 2004).

Commenting on the disagreement between findings from this study and those of previous studies, Baker et al. (2004) asserted that differences in achievement as a function of instructional time only emerged from comparing extremely low amounts of time with some threshold amount, and then a diminishing return would be seen beyond that point. Therefore, schools should not waste resources in marginal increases in instructional time, as long as the system was within world norms. Baker et al. (2004)

further suggested that if schools had a choice between using resources to increase time versus improving teaching and the curriculum then the schools should give priority to the latter.

Yair's study (2000) is another study that opposed the positive relationship between instructional time and student achievement. Using productive time rather than allocated time as an indicator of instructional time, Yair (2000) estimated the effects of productive time on the probability of students' engagement in instruction. Results from logistic regression analysis showed that students were disengaged a large portion of the time in academic classes, and that the existing instructional methods and strategies produced low rates of productive time, especially for minority students. Specifically, African American students reportedly were disengaged from instruction 51% of the time, and Hispanic students 52% of the time. In contrast, Whites and Asian American noted engagement with their lessons 6% to 10% more often than their African American and Hispanic peers. The study also found that student disengagement became more prevalent as students advanced to higher grades.

Not surprisingly, Yair (2000) found that student disengagement was associated with subjects taught and instructional methods and strategies. For example, of all the instructional strategies (i.e., laboratory work, presentation, group work, use of TV and video, individualized instruction, and teacher lectures), teacher lectures, the most prevalent strategies used in the United States schools appeared to attract the lowest rates of student-reported engagement. Based on these findings, Yair (2000) concluded that instructional reforms rather than the simple addition of time would be more productive in raising student achievement and in bringing about greater social equality in education.

## Teacher-related Factors and Student Achievement

What students are expected to learn, how the instruction is organized and delivered, and what students have learned are believed to stem from experiences and values embedded in the professional training and development of teachers (Cogan & Schmidt, 1999). Similarly, Wright, Horn, and Sanders (1997) believed that teachers' effects are dominant factors affecting students' achievement gain. While it is clear that teacher quality play an important role in student learning, there is considerably less consensus on which teacher-related characteristics are strongly related to students' higher performance. The following section focuses on research studies that examined the effects of teacher-related factors such as preparation to teach, readiness to teach, and professional development on students' achievement both in the United States and in the international setting.

### *Preparation to Teach*

A variety of research studies have examined the association between how teachers were prepared to teach and their students' achievement (Bankov, Mikova, & Smith, 2006; Darling-Hammond, 2000; Greenberg, Rhodes, Ye, & Stancavage, 2004; Grouws, Smith, & Sztatjn, 2004). These studies, however, produced mixed results. For example, Ferguson's (1991) analysis of Texas school districts found that teachers' expertise, including their scores on a licensing examination measuring basic skills and teaching knowledge; master's degrees; and experience accounted for more of the inter-district variation in students' reading and mathematics achievement in Grades 1 through 11 than student socioeconomic status. The effects were so strong, and the variations in teacher expertise so great, that after controlling for socioeconomic status, the large disparities in achievement between Black and White students were almost entirely accounted for by differences in the qualifications of their teachers (Ferguson, 1991). Similarly, Darling-Hammond (2000) also examined a study conducted in 1999 by Los Angeles County Office of Education on elementary student reading achievement and found that across all income levels, students' reading achievement was strongly related to the proportions of fully trained and certified teachers, much more so than to the proportion of new teachers in the school. The study concluded that differences in students' test scores was a teacher training issue and not due to new teachers' lack of classroom experience (Darling-Hammond, 2000).

Likewise, evidence from a study of Grouws, Smith, and Sztajn (2004) suggested that teacher's undergraduate major in mathematics appeared to influence eighth grade student mathematics performance on the National Assessment of Educational Progress (NAEP). However, an examination of this effect for fourth grade students yielded non-significant results. The researchers explained that the observed difference in the effect of preparation to teach between eighth grade and fourth grade might be due to the fact that, at the elementary level, teachers were expected to teach different subjects regardless of their undergraduate major field of study. However, at the middle school level, teachers were expected to teach the subject related to their undergraduate field of study. Therefore, it was reasonable to observe a stronger relationship between teachers' preparation to teach and student achievement at the higher educational level than at the lower educational level (Grouws, Smith, & Sztajn, 2004).

Also using the data from NAEP 2000 for eighth-grade math, Greenberg, Rhodes, Ye, and Stancavage (2004) investigated the relationship between teacher qualifications (i.e., certification, academic major or minor, highest degree, total teaching experience and experience teaching mathematics) and student achievement. Multiple regression was employed as a statistical analytic method in this study. In order to estimate the independent effect of each teacher attribute on math achievement, this model controlled for student gender, ethnicity, eligibility for free and reduced lunch program, number of reading materials at home, and parental education. In addition, this study also applied sampling weights, replicate weight and average plausible value as a way to address the issue of complex sample design of NAEP. This research supported the findings from previous studies that teaching certification was positively associated with higher math achievement. With regard to the effect of academic major, students across all math ability levels (i.e., low, medium, and high) who had teachers with a major in math scored higher than their peers whose teachers had a major outside of their field of teaching. In terms of teaching experience, students who had teachers with more than five years of experience teaching math tended to perform better in math than their friends who had less experienced math teachers (Greenberg, Rhodes, Ye, & Stancavage, 2004).

Despite widespread evidence suggesting that preparation to teach enhances student learning, there was a group of researchers who argued that such a conclusion was not always true. Using the example of Teach for America (TFA) teachers, Glazerman, Mayer, and Decker (2006) demonstrated that having no preparation to teach (i.e., not having a college degree in math education, math teaching certification, or math teaching experience) did not prevent TFA teachers from contributing positively to math achievement of their $12^{th}$ grade students. In fact, it was observed that TFA teachers tended to produce significantly higher student test scores than the other teachers in the same schools – not just certified novice teachers but also certified veteran teachers. Glazerman, Mayer, and Decker (2006) concluded that the salient factors of TFA teachers' success in teaching are high academic records in any field of study, motivation, and enthusiasm to teach.

On the international setting, the contention that a positive relationship exists between teachers' preparation to teach and student achievement was also not supported. A study conducted by Bankov, Mikova, and Smith (2006) in Bulgaria is an example. Using HLM to analyze TIMSS 2003 data for eighth grade math and science, this research suggested that having a teacher who had a major or main area of study in the subject taught was not associated with greater math or science achievement. Unexpectedly, students who had a life-science teacher with a degree in biology tend to have lower scores on the life-science assessment than students whose teachers did not have a degree in biology. In explaining the contradicting results observed in this study relative to those of the United States, the researchers stated that "traditional measures of teacher quality, including a match between the subject matter that teachers have studied, their level of experience, and their self-assessment of their readiness to teach their subject material, may not be relevant for Bulgaria at this juncture in its educational reform" (p. 471).

### *Readiness to Teach*

In an effort to define pathways that teachers can follow to become successful in the "theory-rich, open-ended, and content-intensive classrooms", Shulman and Shulman (2004) created a model called "Teacher Learning Communities" (Shulman & Shulman, 2004, p. 259). In this model, readiness to teach is highlighted as one of the five important elements (i.e., ready, willing, able, reflective and communal) that teachers must develop along their paths to be a successful teacher. As Shulman and Shulman (2004) put it: "An accomplished teacher is a member of a professional community who is ready, willing, and able to teach and to learn from his or her teaching experiences" (p.259). According to Shulman and Shulman (2004), a teacher's readiness to teach is determined by the teacher's development of visions of teaching and learning. Specifically, a well-defined vision of teaching and learning (e.g., teaching is a process other than telling, or learning is a process other than repeating or restating) serves as a goal toward which teacher development is directed, as well as a standard against which teachers' thoughts and actions are evaluated (Shulman & Shulman, 2004).

Because teacher's preparedness to teach, self-confidence, and motivation play important roles in shaping a teacher's vision of teaching and learning, it can be inferred that these factors are related to teacher readiness to teach (Bankov, Mikova, & Smith, 2006; Darling-Hammond, 2000; Shulman & Shulman, 2004). Darling-Hammond (2000) asserted that teachers who lack adequate teaching preparation tend to have poor visions of teaching and learning which, in turn, negatively affected teacher readiness to teach and their student outcomes. In illustrating this point of view, Darling-Hammond (2000) cited a teacher, a graduate from Yale University, who did not have an undergraduate major in education but believed that with his intelligence and enthusiasm for teaching he would be able to help his students learn. Unfortunately, this was not the case.

> I – perhaps like most TFAers [Teach for America teachers] – harbored dreams of liberating my students from public school mediocrity and offering them as good an education as I had received. But I was not ready… As bad as it was for me, it was worse for the students. Many of mine … took long steps on the path toward

> dropping out…. I was not a successful teacher and the loss to the students was real and large. (p.168)

Interestingly, Glazeman, Mayer, and Decker (2006) recently conducted a randomized experimental study on the impacts of Teach for America teachers on student achievement and other outcomes and found contradicting results. This study suggested that the allowance of Teach for America teachers to bypass the traditional route to the classrooms did not seem to harm students. In fact, there were statistically significant positive effects of Teach for America teachers on 12th grade students' mathematics achievement. The study, however, found that Teach for America teachers were more likely to report problems with student behaviors than regular teachers who had teaching certificates or undergraduate majors in education.

Bankov, Mikova, and Smith (2006), on the other hand, argued that having a teaching certification or a major in education does not necessarily guarantee that a teacher is ready to teach. With schools around the world becoming increasingly diverse, a teacher who knows the subject matter well but lacks essential understanding of culturally diverse classrooms would be unlikely ready to teach. This is because, as Hollins (1995) points out, any cultural mismatch between the teacher and students can potentially interfere with instruction and learning.

Building on the similar point of view that teacher readiness to teach is greatly dependent on their understanding of a culturally diverse classroom, Wiggins and Follo (1999) emphasized the importance of teacher motivation and willingness to learn of others' cultural differences. Wiggins and Follo (1999) contended that despite prior exposure to a culturally diverse environment, a teacher who is not willing to learn of others' cultural differences is unlikely to achieve a desirable understanding level needed to foster his/her readiness to teach. It is important to note, however, that although this contention is sound from a theoretical perspective, there is insufficient empirical evidence to support this view. Future studies, therefore, should pay more attention to this line of research, wherever possible.

### *Professional Development*

Teacher professional development is instrumental in educational reform efforts to improve student learning (Borko, 2004). According to Jacob and Lefgren (2004), professional development is a common practice in the United States public schools. A study of Parsad et al. (2001) suggested that approximately 72% of teachers reported having participated in training related to the subject area of their main teaching assignment during the previous 12 months. However, despite the widespread implementation of teacher training programs across the country, research linking teacher professional development with student performance is inconclusive (Jacob & Lefgren, 2004; Johnson, Kahle, & Fargo, 2007; Ross, Bruce, & Hogaboam-Gray, 2006 ).

In a meta-analysis study, Kennedy (1998) reviewed 93 studies but found positive effects of teacher development on student achievement in only 12. In line with these findings, Corconran (1995) and Little (1993) claimed that typically teacher professional development programs are low-intensity activities that lack continuity and accountability. More than half of the teachers surveyed reported engaging in only eight hours or less of training per content area per year (Corconran, 1995). Holding a similar view, Borko (2004) criticized existing professional development programs for their failure to take into account how teachers learn. For example, in a seminar on community learning, teachers were expected to create a community of learners among their students but the teachers themselves were not provided a parallel community in the training to nourish their own growth. As a consequence, upon completion of these programs, many teachers still felt they were not ready to teach. Borko (2004), therefore, concluded that these programs were "woefully inadequate" and "intellectually superficial" (p. 3). Likewise, Ross, Bruce and Hogaboam-Gray (2006) questioned findings of several studies on professional development, arguing that all of them were deficient in some way. For example, with the study of Hamilton et al. (2003), it was impossible to extract the unique contribution of professional development on student outcomes because the study did not control for the provision of innovative curriculum materials which could account for the student achievement (Ross, Bruce, & Hogaboam-Gray, 2006). Similarly, for the study of Reys et al. (1997), the findings were biased because they were based on an unrepresentative sample of teachers (i.e., 80% had masters' degree and 40% were members of National Council of Teachers of Mathematics) (Ross, Bruce, & Hogaboam-Gray, 2006).

Opposing the assertion that teacher professional development programs did not yield improved instructional practices and student learning, Smith and Neale (1991) conducted a study to examine the impact of the Cognitively Guided Instruction (CGI) project. The CGI project is a 4-week summer workshop which aimed to increase teachers' ability to explore student thinking and to plan ways to build on students' knowledge in math instruction. In this project, participating teachers were randomly assigned into two groups, treatment and control. This project showed that, by the end of the workshop, teachers in the treatment group reported an increased awareness of the role that children's thinking plays in the learning process, and the importance of listening carefully to students in order to build on their understanding and misconceptions (Smith & Neale, 1991). Specifically, in comparison with the teachers in the control group, CGI teachers appeared to know more about the strategies that children use to solve problems, the kinds of problems they find difficult, and different ways to pose problems to students (Smith & Neale, 1991). In regard to student learning, the study found that, during the year following the summer workshop, students in the CGI classrooms solved a wider variety of math problems, used more problem-solving strategies, and were more confident in their math ability than were students in control classrooms (Carpenter & Fennema, 1992).

Consistent with this finding, Johnson, Kahle, and Fargo (2007) recently conducted a quasi-experimental study to examine the effect of sustained, whole-school professional development on sixth to eighth grade student achievement in science. In this 3-year (2002-2005) longitudinal study, science achievement of students of 11 science teachers from a treatment school was compared with that of students of six science teachers from a control school. Each school had approximately 750-900 students. At the treatment school, science teachers were offered an intensive 80-hour professional development program during the summer of the first year, followed by 36 hours across each of the three academic years, for a total of 198 hours. The training emphasized standards-based instructional practices (i.e., instructional strategies focusing on inquiry as central mode for teaching science). For this study, a cross-sectional multiple regression analysis that adjusted for cluster sampling was conducted for each year. Results of the study showed that there was a positive relationship between student science achievement and teacher participation in professional development program. Specifically, students' repeated involvement in improved instruction resulted in significant achievement gains by both majority and minority students in Year 2 and 3. For Year 1, there was no significant difference in achievement scores between the two study groups of students. This study suggests that duration of professional development is linked to increased student achievement scores. This is somewhat expected because the more opportunities teachers have to practice their newly learned skills, the deeper and more sustained their experiences became which, in turn, positively influence student learning (Johnson, Kahle, & Fargo, 2007).

According to Loucks-Horsley, Hewson, Love, and Stiles (1998), there are five types of teacher professional development: immersion, examining practice, curriculum development, curriculum implementation, and collaborative work. The first type, immersion strategies involve having teachers actually "do" science or mathematics and gain the experience of doing science or math with a scientist or mathematician. The second type, curriculum implementation involves having teachers using and refining the use of instructional materials in the classroom. The third type, curriculum development involves having teachers help create new instructional materials to better meet the needs of students. The fourth type, examining practice includes case discussion of classroom scenarios or examining real classroom instruction. And finally, the fifth type, collaborative work includes study groups, peer coaching; mentoring and classroom observation and feedback.

Interested in how eighth-grade science and math achievement was associated with the types of professional development (i.e., immersion, examining practice, curriculum development, curriculum implementation, and collaborative work), Huffman, Thomas, and Lawrenz (2003) analyzed the data that were collected from 94 science teachers and 104 mathematics teachers in 46 schools across a southern state of the United States. The dependent variables were student achievement scores from the state standardized tests in math and science and the independent variables were the five types of professional developments. Results from regression analyses suggested that there was only a weak relationship between these types of professional development and student achievement on state exams. Specifically, only curriculum development for math teachers was found to relate to student math achievement; however, the relationship was negative. None of the different types of professional development were significantly related to student science achievement. Mathematics teachers with students who have lower achievement were found to engage in more long-term curriculum development. In this study, curriculum development for math teachers accounted for 16% of the variance of student achievement.

Using the data from the California Learning Assessment System (CLAS) 1993, Wiley and Yoon (1995) examined the relationships between student math achievement and teacher-related variables such as familiarity with curriculum goals and standards, participation in professional development, and implementation of instructional practices. For this study, approximately 1,750 math teachers (1,100, 420, and 230 teachers in Grades 4, 8, and 10, respectively) and 30,250 students (17,250, 10,100, and 3,000 students in Grades 4, 8, and 10, respectively) were included. Findings from this study suggested that, in Grade 4, for those teachers who were familiar with mathematics instruction assessment guides, and who participated in mathematics curriculum activities, students performed significantly better than students of the teachers who were not involved in those activities. Interestingly, participation in professional conferences formed an exception to this pattern. A similar pattern was also observed in Grade 8. Grade 10, however, showed the least impact despite the highest level of teachers' familiarity with math goals and standards and frequent participation in various instructional activities.

Similarly, Cohen and Hill (1998) conducted an experimental study to examine the extent to which student math achievement was associated with teacher participation in professional development programs that focused on teaching mathematics content in the state of California. Findings of this study showed that, after adjusting for student background variables, experimental schools where teachers participated in professional development programs had significantly higher average math achievement than control schools where teachers did not participate in this type of professional development. Kennedy (1998), however, found that students whose teachers participated in specific content-related professional developments showed better conceptual understanding in math and science than their peers whose teachers only participated in general professional development programs.

## School-related Factors and Student Achievement

School systems around the world differ in many respects. Existing literature has identified many potential factors that can bear upon the differences in student achievement between and within schools. Examples of these factors include availability of school resources, differences of student backgrounds and characteristics, teacher quality, class size, and instructional time to name a few. Given these differences, many researchers have argued that findings from one school system should not be compared with or generalized to other school systems, and therefore, researchers should focus their attention to research issues within individual school systems. This point of view, however, could not stand with time without criticism. There were a group of researchers who strongly believed that despite sizeable differences across school systems, every school system can greatly benefit from other systems by carefully examining and interpreting their findings. In so doing, each school system would learn more about themselves -- where they are relative to other school systems and most important of all, what opportunities as well as threats they should consider in the quest to improve their student achievement. Thus, it would be very worthwhile for educational researchers to pursue this line of research. In this particular section, the focus of the literature review will be on research studies that investigated the effects of school-related factors such as class size, availability of school resources and instructional limitation on student achievement across countries. These research topics were selected because they have been hotly debated in the United States for many years and have recently expanded to other countries (Luyten et al., 2005). In addition, a better understanding of these topics would likely result in better application of research findings to school policy and practices.

### *Class Size*

There is ample literature on the relationship between class size and student learning. The results of this work, however, are varied. In 1978, Glass and Smith published results from their meta-analysis of 77 studies that investigated the effects of class sizes on student achievement (Glass & Smith, 1978). Several important findings were drawn from this study: (a) overall, small class sizes were associated with higher student achievement, (b) the effects of class size appeared to grow as size was reduced, meaning a reduction from 10 to 5 students had a greater impact than a reduction from 30 to 25 students, and (c) the relation between class size and achievement was similar across students of different ages and ability levels (Glass & Smith, 1978).

Interestingly, in the same year, the Educational Research Service (ERS, 1978) conducted a review of 41 studies to examine the relationship of class size and student achievement across grade levels. Major conclusions from this study are quite different: (a) the relationship between class size and student achievement was highly complex, (b) the effects of class sizes were a product of many variables, including subject areas, student characteristics, learning objectives, class and school resources, and teacher qualities, (c) within the mid-range of 25 to 34 students, class size appeared to have little impact on achievement of students in the primary grades, and (d) small class size appeared to be most beneficial for students with either lower academic ability or economically or socially disadvantaged backgrounds (Cooper, 1989b). In responding to the dissimilarities among the results of two studies, ERS (1980) criticized Glass and Smith (1978) for obscuring important distinctions in class size research and for over generalizing their major findings which were based on too few studies. As a conclusion, ERS (1980) expressed critical need for further research in this area.

Disagreeing with both sets of results produced by Glass and Smith (1978) and ERS (1978, 1980) because “neither adequately considers the quality of the critical evidence” (Slavin, 1989, p. 102), Slavin (1989) conducted another meta-analysis study of the effects of class size on achievement. Unlike prior research, Slavin imposed several restrictions to his study such as (a) achievement scores had to be standardized scores; (b) large classes had to be compared to classes that were at least 30% smaller and contained no more than 20 students, and (c) the study had to use random assignment to alternative class sizes. As a result, this research suggested that reducing class size would not in itself make a substantial difference in student achievement even at the lower grades. Similarly, reducing class size was not likely to solve the achievement problems of at-risk student unless the class size was reduced to one student per class (Slavin, 1989).

As one of the most ambitious experiments ever attempted in the United States education, Project STAR, a longitudinal study (1985-1989) investigated the effects of class size on math and reading achievement of 6,829 kindergartens to third grade students in Tennessee who were randomly assigned to small classes (13-17 students) and large classes (22-26 students) (Pong & Pallas, 2001). In convergence to the results of ERS (1978), findings from Project STAR suggested that small classes tended to increase student math performance in the early grades by about one third of a standard deviation (Pong & Pallas, 2001). Also, a follow-up experiment of the students participating in the Project STAR revealed that the benefits of small classes persisted significantly for six years after the students returned to regular-sized classes at ninth grade (Nye, Hedges, & Konstantopoulos, 2001). Additionally, in examining whether the effects of class size functioned different across students of different backgrounds, Nye, Hedges, and Konstantopoulos (2001) found that the lasting effects of smaller classes were greater for minority students than for White students.

Recently, with international educational data becoming more accessible, many researchers have taken this opportunity to investigate the effects of class size on achievement across countries. Not surprisingly, results from these studies were also mixed. As indicated in the study of Pong and Pallas (2001), class sizes in Asian countries tended to be quite large by the United States standards. For example, data from the TIMSS 1995 suggested that about 60% of Hong Kong's classes had an average of 39-42 students. Similarly, in Korea, the majority of the classes (64%) had an average of 50-54 students. However, students in these countries consistently scored at the top in international math achievement tests. An examination of class sizes within countries also yielded similar results. That is, high performance classes tended to be larger than average classes. After adjusting for spurious factors in HLM models for individual countries, Australia and Canada were the only non-Asian countries where larger classes led to better performance in math than did smaller classes. By contrast, in the United States, small classes with fewer than 19 students outperformed their large class counterparts (Pong & Pallas, 2001).

Also using the TIMSS 1995 data for math and science, Woobmann and West (2006) estimated the effects of class size on seventh and eighth grade student performance in 11 countries (i.e., Belgium, Canada, Czech Republic, France, Greece, Iceland, Portugal, Romania, Singapore, Slovenia, and Spain). This study found sizeable beneficial effects of small classes in Greece and Iceland, but not in other countries. It is important to note, however, that in both Greece and Iceland, students tended to perform below the international average whereas in the remaining countries, where the effects of class size were statistically non-significant, students tended to perform above the international average. These results were interpreted by Woobmann and West (2006) to mean that within their own educational systems (i.e., Greece and Iceland) class-size reduction seemed to be associated with improvement of student achievement. However, in consideration of a larger picture, students in small classes in Greece and Iceland did not perform as well as students in large classes of other countries included in the study (Woobmann & West, 2006).

### *School Resources*

Research on school effects on student achievement has a fairly long history, dating back since 1966 when the seminal work of Coleman and associates on the relationship of school effects relative to family effects on student achievement was published. In this report, Coleman et al. (1966) showed that, in the United States, compared with family-related factors, school-related factors had only modest effects on student achievement (Suter, 2000; Baker et al., 2002). Later, in 1972, Mosteller and Moynihan conducted a re-analysis of Coleman's study and found similar results. That is, school variance had little influence on student achievement (Mosteller & Moynihan, 1972).

Challenging these findings, Comber and Keeves (1973) conducted a much larger study which included 19 countries in the world, including the United States. The data came from the IEA's (1971) First International Science Study (FISS). Evidence from this study suggested that school quality (i.e., instructional practices and instructional resources) was directly related to science achievement in middle and high schools in participating countries. In questioning such contradicting results, Coleman (1975) conducted a re-analysis of the Comber and Keeves' (1973) study using a different research design and statistical data analysis method. Despite several differences in the obtained results compared to those of Comber and Keeves' (1973), Coleman (1975) concurred that school variables had significant effects on 10- and 14-year old students' achievement in science in the six countries he studied, including the U.S.

More empirical evidence of school effects on student achievement was reported by Heyneman and Loxley (1982) when they re-analyzed the same IEA data for science education in 19 countries that were analyzed earlier by Comber and Keeves (1973) and Coleman (1975). Findings from this re-analysis study were important in that they not only confirmed the significance of the effects of schools on student achievement but more importantly, they suggested that in some developing countries, school effects could outweigh the effects of home background. For example, in India, the effects of school and teacher quality could account for up to 90% of the variance in student achievement (Heyneman & Loxley, 1982).

In an attempt to better understand the variation of school effects on student achievement in developing countries, Fuller (1987) examined a series of studies and concluded that after accounting for the effect of student background, schools exerted a greater influence on achievement of students in developing countries than in developed countries. Three reasons were used to explain the findings. First, due to lack of available material resources at home and at schools in developing countries, the influence of social practices within classrooms may play a greater role than do material inputs, as appeared to be the case in the United States. Second, social class structures in developing countries often are less differentiated than in highly industrialized societies. Thus, advantages rooted in social class and related parenting practices tended to be less influential in developing countries. Lastly, the school institution often operates within communities where any commitment to written literacy or numeracy is a historically recent event.

Therefore, a school of even modest quality may significantly influence academic achievement (Fuller, 1987).

More recently, with an interest in finding out whether the effect of national development on the association among family SES, school resource quality, and achievement found in data from the 1970s were still evident in the mid-1990s, Baker, Goesling, and LeTendre (2002) analyzed the TIMSS 1995 data for 36 countries. In this study, national economic development was defined using World Bank's (1994) index of gross domestic product (GDP) per capita. Family SES was a composite score of mother's and father's education level and number of books in the home. School resource quality was a composite score representing availability of 11 indicators: instructional material, budget for supplies, school building space, heating and lighting, instructional space, computer hardware, computer software, calculators, library materials, audiovisual resources, and library equipment. In analyzing the data, Baker et al. (2002) applied a more advanced statistical analysis method (i.e., hierarchical linear modeling) to account for the dependence of nesting data of the TIMSS 1995. Interestingly, results from this study indicated that the relative effect of school resources and family background on achievement within nations was no longer associated with national income levels in the way originally described in the studies of 1970s. Specifically, low-income nations did not show stronger school effects than high-income nations. However, across nations, it was evident that low-income countries tend have low achievement scores and high-income countries tend to have high achievement scores (Baker et al., 2002)

Similarly, Wobmann (2003) used TIMSS 1995 data from 39 countries with 260,000 middle school students to investigate the impact of differences in schooling resources and educational institutions on student performance. The data were analyzed using hierarchical multilevel modeling (HLM). Missing data in the study were handled by imputation. If missing values came from a discrete variable ordinary least square (OLS) estimation was used. If missing data came from a binary variable, a probit model was employed. And, if missing data were from a polytomous variable, an ordered-probit model was applied. The results from student-level estimation suggested that international differences in student performance could not be attributed to school resource differences but were considerably related to institutional differences. In particular, institutional combined factors that yielded positive effects on student achievement included centralized examinations and control mechanisms, school autonomy, individual teacher influence over teaching methods, limits to teacher unions' influence on curriculum scope, scrutiny of students' achievement and competition from private schools (Wobmann, 2003).

### *Instructional Limitations*

Teachers' perception of the extent to which their instruction is limited due to student factors such as unwillingness of student to learn, heterogeneity of student background (e.g., family SES, language, special needs) and differences in student academic levels is related to teacher's efficacy, teacher confidence, and teacher flexibility. According to Tschannen-Moran and Hoy (2001), "A teacher's efficacy belief is a judgment of his or her capabilities to bring about desired outcomes of student engagement and learning, even among those students who may be difficult or unmotivated" (p. 783). Because efficacy affects the effort teachers invest in teaching, the goals they set, and their level of aspiration, teacher efficacy, in theory, is related to student achievement (Tschannen-Moran & Hoy, 2001). Research studies that lend support to this position are numerous. For example, Guskey (1987) indicated that teachers with a strong sense of efficacy tend to exhibit greater levels of willingness to experiment with new methods to better meet the needs of their students. Similarly, several researchers found that, in the face of setbacks, efficacy beliefs tend to influence teachers' instructional approaches in a myriad of positive ways: being less critical of students when they make errors (Ashton & Webb, 1986), working longer with a student who is struggling (Gibson & Dembo, 1984), exhibiting greater enthusiasm for teaching (Allinder, 1994), and having greater commitment to teaching (Coladarci, 1992).

Believing in the powerful effect of teacher efficacy on student achievement, Glazerman, Mayer, and Decker (2006) conducted a randomized experiment study to compare the effects of teacher efficacy with the effects of teaching qualification and teaching experience. In this study, efficacious teachers were defined as Teach for America (TFA) teachers. TFA teachers were recruited by TFA program and assigned to teach in schools that serve a disadvantaged, largely minority population of students. In general, TFA teachers were recent graduates of the nation's top colleges with strong academic records (GPA = 3.5 and above). They were described as enthusiastic and committed to teaching even though they did not have education-related majors in colleges or any student teaching experience. The teachers in the comparative group consisted of regular teachers recruited by schools who varied in both teaching qualifications (i.e., certified and uncertified) and teaching experience (i.e., beginning teachers and teachers with more than 5 years of teaching experience). Results from this study suggested that students who had efficacious teachers scored significantly higher in $12^{th}$ grade math than their peers who had regular teachers. The size of the impact was relatively large, corresponding to about 10 percent of grade equivalent or an additional month of instruction. More importantly, this finding was consistent across all the subgroups and regions included in this study (Baltimore, Chicago, Los Angeles, Houston, New Orleans, and the Mississippi Delta). The researchers concluded that the effects of teacher efficacy outweighed the effects of teacher qualification as well as teaching experience. It is important to note, however, that in this study the researchers did not use any official instruments to assess the level of teacher efficacy in TFA teachers; rather the researchers assumed TFA teachers were efficacious teachers, basing on the definitions of TFA program.

With regard to teacher self-confidence, substantial evidence also indicates that teachers with a high level of confidence tend to be successful with students (Darling-Hammond, 2000). Teacher self-confidence in this context is not restricted to teacher adept knowledge in the subject matter; rather teacher self-confidence reflects teacher competence in multiple skills areas such as knowing how to convey the material in different ways that can benefit students with various learning ability, knowing how to manage the classroom so that a sufficient amount of time can be devoted for instruction, and knowing how to use different methods to assess student learning.

With schools promising to serve a much more diverse group of students to much higher standards, teacher flexibility in teaching has become an important quality to warrant effective teaching and learning in school (Darling-Hammond, 2000). Flexibility allows teachers to move beyond their own cultural boundary, to put themselves in the shoes of the students who are quite different from them, and to adapt instruction to students' individual learning needs (Darling-Hammond, 2000).

As educational reform movements in the United States and around the world are setting ambitious goals for student learning, Borko (2004) believed that flexibility would help teachers to better implement the changes in the classroom practices demanded by these reforms. To elaborate, Berko (2004) added that because the magnitude of these changes can be large, a great deal of learning would be required on the part of teachers. Therefore, without flexibility of an open mind, it would be difficult for the teachers to make the changes. This position is also supported by evidence from the study of Carpenter et al. (2004) that teachers who had a lower level of flexibility tended to ignore new instructional practices required by the math reform if these practices conflicted with their views of mathematics teaching. Similarly, building on this point of view, Borko and Putnam (1996) posited that, to foster students' conceptual understanding, teachers must have rich and flexible knowledge of the subject they teach. That is, in addition to the essential knowledge of the discipline, teachers must be able to use multiple ways to connect ideas and organize learning processes so as to help students construct new knowledge.

Examining teacher flexibility from an international perspective, Bryan, Wang, Perry, Wong, and Cai (2007) found that overall teachers from the four countries studied (Australia, United States, Mainland China, and Hong Kong) shared the similar view regarding the importance of being flexible in teaching in order to meet students' needs. As one teacher from Hong Kong stated "The teacher should not just blindly follow the lesson plan and let the lesson go on without considering students' response" (p. 336).

Similarly, a teacher from the United States affirmed "Being able to observe, and judge, and evaluate each student and meeting their individual needs probably is the most difficult and probably one of the most crucial parts for an effective teacher" (p. 336). The teachers from Mainland China and Hong Kong, however, cautioned that in these countries, teacher's flexibility sometimes was compromised due to the large number of students in the class and the amount of content that is required to be covered in a lesson (Bryan et al., 2007).

Based on these research findings, it can be inferred that, given a class of students with various backgrounds and characteristics, those teachers who possess a high level of efficacy, self-confidence, and flexibility are less likely to view the class as limitation to instruction; rather they tend to exert a higher commitment to improve student learning through application of customized instruction to meet their student needs. Therefore, from the school perspective, schools would seem to be in a better position to improve their students' achievement level if they have more teachers with a high level of efficacy, self-confidence, and flexibility. One solution for schools to achieve this goal is to improve teacher efficacy, self-confidence and flexibility by providing teachers with regular professional development programs. Schools should also emphasize these qualities when recruiting new teachers.

## Summary

Student mathematics achievement at the national level has often been associated with the future economic power and security of a country. Thus, the desire to understand and identify factors that are related to increased student mathematics achievement has become a national goal in many countries around the world, including the United States.

Over the years, numerous studies have been conducted across countries to investigate the effects of contextual factors on student mathematics achievement. These contextual factors include but are not limited to student background, instructional practices, teacher background, and school background variables. Findings from these studies, however, have shared little consensus.

The purpose of this chapter was to provide a comprehensive review of existing literature on the relationship between student mathematics achievement in middle school and the aforementioned contextual factors from a national as well as an international perspective. In order to allow for a broader inclusion of empirical studies, this chapter also reviewed research that examined student achievement in subject areas other than mathematics, such as science, reading, literacy, and civics across different grade levels.

Through this comprehensive review of literature, several important findings can be highlighted. First, in terms of variable operationalization, it seems common that achievement outcomes were reported in the form of standardized achievement scores. These standardized scores came from various data sources, including international achievement assessments, national achievement assessments, state achievement assessments, and local achievement assessments. With regards to operationalization of achievement outcomes, measures of student achievement tended to be fairly consistent. For example, there appear two common definitions for math achievement: (a) math as an average composite score of sub-content areas such as number, data, algebra, measurement, and geometry, and (b) math as a single sub-domain score such as algebra, or measurement, or problem solving. In contrast, for contextual factors, the opposite seems true. Background variables were defined variously from one study to another. In addition, the majority of the background data came from self-reported questionnaires. For example, in some studies, time on homework referred to time student spent on homework in all subjects per week; whereas in other studies, this variable was defined as time student spent on mathematics homework per week. Still, in other studies, time on homework was operationalized as the time students typically spent on homework per day. Yet, in other studies, time on homework was an aggregated variable which consisted of homework frequency and homework length. Thus, it is essential for research consumers to interpret research findings with caution due to differences in measure operationalizations.

Second, there seems little consensus among the studies reviewed in this chapter regarding potential contextual factors that could improve student math achievement. In addition, the strength and direction of the relationships between contextual factors and math achievement appeared to be inconsistent from one study to another. For example, whereas Coleman (1966) suggested that family-related factors exerted stronger positive effects than school-related factors on student achievement, Komber and Keeves (1973), Heyneman and Loxley (1982), and Fuller (1987) argued the opposite was true, especially if the studies were conducted in developing countries. Similarly, the relationship between time on homework and achievement has been hotly debated. Cooper (1989) found that time student spent on homework was positively related to greater student achievement. In contrast, evidence from Rodriguez's (2004) study suggested that students who did no homework each day performed slightly higher on average than those students who spent more than one hour a day on math homework. In addition, Trautwein (2007) indicated that the relationship between homework time and achievement was only moderate at the school level and was negative at the student level. Likewise, whereas Darling-Hammond (2000) strongly believed that teacher preparation to teach was one of the most important determinants of student achievement, Glazerman, Mayer, and Decker (2006) demonstrated that having no preparation to teach did not prevent talented and enthusiastic individuals who neither majored in education nor had prior teaching experiences from contributing positively to student achievement. Despite discrepancies in research findings, all of the researchers acknowledged that their studies were limited to certain extents and thus, their findings should be carefully interpreted within context. Similarly, all the researchers agreed that further research was needed in order to provide more evidence regarding the relationship between contextual factors and student achievement.

Third, the majority of studies reviewed in this chapter were guided by a correlational research design because the focus of these studies was the relationships between contextual factors and student achievement. However, these studies differed from one to another in several respects, including data sources, samples selection, variables of interest, data management (e.g., treatment of missing data and use of sample weight to account for complex, large-scale survey design), and methods of data analysis. As an illustration, with regards to variables of interest, different approaches were used to identify the final set of variables for the study. In some studies, the researchers determined potential predictors of student achievement by examining prior research. In other studies, only variables that met certain statistical significance criteria (e.g., correlation coefficients of the variable with achievement had to be larger than twice their standard errors or average standardized regression coefficients across samples exceeded .05) could be selected. Still in some studies, variables of interest were only included if they had sufficient data across samples. Yet, in other studies, the list of variables could be different across samples.

In terms of data analysis, several common statistical methods such as multiple regression analysis, structural equation modeling (SEM), hierarchical linear modeling (HLM) were employed to examine the effects of contextual factors on student achievement across studies. For studies conducted before the 1990s, multiple regression analysis and SEM tended to be used more frequently and for studies conducted after 1990s, HLM appeared to be used more frequently. The reason behind this shift in the method of data analysis over time was due to the fact that HLM is a newer and more advanced statistical analysis method that allows researchers to conceptualize the effects of contextual factors on students achievement as occurring at multiple levels due to the nature of nested structure of educational data (i.e., students nested within teachers, and teachers nested within schools, etc.). Other statistical tests such as independent t-test, analysis of variance (ANOVA), and multivariate analysis of variance (MANOVA) were also used to examine differences across sub-samples (e.g., gender groups, grade levels, and regions) and across time.

There were also several studies that applied qualitative approaches to analyze the data. The area of research that seemed to attract more qualitative studies includes opportunity to learn in terms of curriculum coverage, teacher quality (i.e., preparation to teach, ready to teach, and professional development) and instructional activities in the classroom. Thus, the data sources used in these studies came from interviews with study participants, classroom observations, instructional goals and curriculum, and participants' reflection journals or field notes.

Fourth, there appears bias in the inclusion of countries in international research studies that examine the relationships of contextual factors and student achievement. Specifically, these research studies tended to focus more on developed countries than on developing countries. Countries that were frequently included in international research studies include the United States, Canada, Germany, Japan, Korea, and Hong Kong. The lack of representation of developing countries in international research is not desirable because it is very possible that the relationships between contextual factors and student achievement that were significant in these countries may not be significant in developing countries due to substantial differences in country economic status. In fact this contention was well supported by the study of Fuller (1987) where the researcher showed that the inclusion of developing countries such as India, Chile, or South Africa actually changed the strength and direction of school-related factors on student achievement from little or non-existent (Coleman, 1966) to a strong and positive relationship (Fuller, 1987). Similarly, as Werf, Creemers, Jong and Klaver (2000) suggested, in Western countries, large differences in student achievement were noted between students from different socioeconomic backgrounds. However, in developing countries such differences were much smaller. Thus, it might not be realistic for developed countries to set a goal to improve achievement level of all students with different SES background but it is possible for developing countries to aim to improve the achievement level of all the students in their educational systems.

Finally, from this examination of literature, the importance of continuing research in the area of international achievement assessment is clear. Repeatedly, international studies demonstrated that they significantly contributed to the advancement of the field of educational research by challenging existing beliefs and research findings, by illuminating new ideas and insights into how to improve educational systems, both from a theoretical and methodological perspective, and by offering a host of opportunities for educators, researchers as well as policy makers around the world to share and learn from each other's experience and expertise. Theoretically, results from these international studies are important in that they provide insights into the extent to which the effects of contextual factors such as family resources and school resources on student achievement could change over time or function differently across educational systems (Baker et al., 2002; Coleman, 1966; Heyneman & Loxley, 1982, 1983; Woobman, 2003). This is because the effects of many of the contextual factors are influenced by national economic status which, in turn, is subject to change across time. Methodologically, the large-scale of international data as well as the level of variance among countries provides excellent advancement opportunities for new and improved statistical methodologies to be developed and tested.

# CHAPTER THREE

# METHOD

Chapter Three is organized into the following major sections: the purpose, research questions, research design, and data analysis.

## Purpose of the Study

The purpose of this study was to investigate correlates of eighth-grade students' math achievement in TIMSS 2003 in four countries. Specifically, within each of the countries included in the study, a series of two-level models was constructed using contextual and background factors at both the student and the classroom/teacher/school levels to account for the variance in eighth-grade students' math achievement within and between schools.

## Research Questions

This study was driven by the following set of research questions:

1) To what extent are student background variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, and tutoring in math) associated with TIMSS 2003 eighth-grade math scores in each country?
2) To what extent are home resources variables (i.e., availability of calculator, dictionary, computer, and desk for student use) associated with TIMSS 2003 eighth-grade math scores in each country?

3) To what extent are instructional variables (i.e., opportunity to learn, activities in math lessons, amount of homework assignment, and instructional time) associated with TIMSS 2003 eighth-grade math scores in each country?
4) To what extent are teacher-related variables (i.e., preparation to teach, ready to teach, and professional development) associated with TIMSS 2003 eighth-grade math scores in each country?
5) To what extent are school-related variables (i.e., class size, school resources for math instruction, and math instructional limitation) associated with TIMSS 2003 eighth-grade math scores in each country?

## Research Design

### *Data Source*

Data from the Trends in International Mathematics and Science Study (TIMSS) 2003, a study conducted by the International Associations for the Evaluation of Educational Achievement (IEA) and maintained by the National Center for Education Statistics (NCES), were used in this study. The TIMSS 2003 database comprised student achievement data in mathematics and science as well as student, teacher, school, and curricular background data for 48 countries at eighth grade and 26 countries at fourth grade (Martin, 2005). For this study, the following databases from TIMSS 2003 for eighth-grade were used: student math achievement, student background, math teacher background, and school background.

The TIMSS 2003 database for eighth-grade math was selected for this study because of its influence on education in both the U.S. and other countries has increased rapidly (Baker & LeTendre, 2005; O'Leary, 2002, Rodriguez, 2004; TIMSS, 1995, 1999, 2003). TIMSS was originally conducted in 1995 and continued every four years, in 1999, 2003, and most recently in 2007. The ultimate goal of TIMSS was to provide trend data on students' math and science achievement from an international perspective (TIMSS, 2003). In 2007, there were more than 60 countries participating in the TIMSS study (TIMSS, 2007). Until the 2007 data are released, the TIMSS 2003 database had the largest and most recent international student achievement data in mathematics and science (TIMSS, 2003). In addition, the TIMSS 2003 database for eighth-grade math included rich and timely information about student, curriculum, teacher, and school background that could be used to examine the relationship between contextual and background factors and student math achievement within and across countries.

### *Sampling Procedures*

TIMSS 2003 used a two-stage sampling design to select representative samples of students in each country. At the first stage, at least 150 schools were randomly sampled. However, because the school sample was designed to optimize the student sample rather than provide an optimal sample of schools, large schools tended to occur in the sample more frequently than in the school population. This sampling design was also known as probabilities proportional to size (Martin, 2005).

At the second stage, one class was randomly sampled in each school. This resulted in a sample size of at least 4,000 students per country. The selection of teachers and school principals was determined by the selection of students because they were linked to the students (Martin, 2005). However, countries could, with prior approval, adapt the sampling design to local circumstances. For example, countries could incorporate in their sampling design important reporting variables (e.g., urbanicity or school type) as stratification variables. At the second stage, countries could also randomly sample one or two classes in each of their schools. Some countries took advantage of this option and sampled more schools and classes resulting in larger sample sizes (Martin, 2005).

### *Data Collection*

The operations of data collection in TIMSS 2003 were on two schedules according to participating countries located in the northern and southern hemispheres. In countries in the southern hemisphere, where the school year typically ends in November or December, the assessment was conducted in October or November 2002. In countries in the northern hemisphere, where the school year typically ends in June, the assessment was conducted in April, May or June 2003 (Martin, 2005).

Each participating country was in charge of carrying out data collection and maintaining quality control procedures. Training manuals for data collection were created for test administrators, detailing standardized procedures such as test security, timing and rules for answering students' questions. In addition, each country nominated one or more persons such as retired school teachers to serve as quality control monitors. These monitors were provided with two-day training sessions (Martin, 2005).

### *Sample*

For this study, two developed and two developing countries were selected from the TIMSS 2003 eighth-grade math database. Two criteria were applied in the sample selection. First, all the countries in the TIMSS 2003 database were stratified into two categories: developed countries and developing countries. The World Bank's (2003) world development indicators were used to classify countries that participated in the

TIMSS 2003 assessment into two categories: developing countries and developed countries (The World Bank, 2003). In 2003, the World Bank used more than 500 indicators to measure development outcomes in 152 countries. Some examples of the World Bank's (2003) world development indicators include population dynamics, labor force structure, employment, national growth patterns, structure of trade, private sector development, investment climate, business environment, stock markets, financial efficiency, and integrated global economy among other specific and quantified targets for reducing poverty, and achieving progress in health, education, and the use of environmental resources (The World Bank, 2003). This step was done to ensure that the study included representative samples of both developed and developing countries. A list of TIMSS 2003 countries grouped by country status (i.e., developing and developed countries) can be found in Appendix A. Next, for each category, the two countries with the largest number of schools were selected. This step was done to ensure that the study has sufficient level two units (i.e., number of schools) for examining the variance in eighth-grade math achievement within each country. Table 1 provides a list of countries that met the selection criterion and their sample descriptions.

It is important to note, however, that due to the complexity of the TIMSS 2003 database, several decisions regarding data management were made during the sample selection process, resulting in some reduction of both level-1 units (i.e., students) and level-2 units (i.e., schools) in the selected countries. For example, whereas in most of the countries, only one math classroom was randomly sampled to participate in the TIMSS 2003, in a few countries, more than one classroom was selected. Similarly, whereas most of the schools had more than 10 students per classroom that participated in the TIMSS

2003, a few schools had less than 10 students. In order to maintain a similar data structure across countries, it was determined that those schools with less than 10 students were removed. In addition, for those schools with more than one classroom sampled, only the classroom with the most number of students was kept for subsequent analyses. Furthermore, because only one class was selected for each school, the number of math teachers was equal to the number of schools.

Table 1.
*Summary of the Samples Included in the Study*

| Country status | Country name | No. of students | | | No. of schools/ teachers | Class size | | |
|---|---|---|---|---|---|---|---|---|
| | | Total | Female | Male | | M | Min | Max |
| Developed | Canada | 8,473 | 4,287 | 4,186 | 354 | 24 | 10 | 38 |
| | U.S.A | 8,008 | 4,106 | 3,902 | 241 | 33 | 10 | 102 |
| Developing | Egypt | 7,095 | 3,329 | 3,765 | 217 | 33 | 12 | 36 |
| | South Africa | 8,927 | 4,470 | 4,355 | 253 | 35 | 10 | 56 |

Note: 102 students in South Africa and 1 student in Egypt did not report their gender

## *Country Profiles*

In an effort to better understand the countries that were selected for this study, a brief profile was created for each of these countries. The types of information that appeared relevant to this study included geographic location and size, population, ethnic groups, languages, political system, roles of government in schools, economic systems, and finally, educational issues. It is important to note that the information in this profile was drawn from multiple sources and priorities were given to the information that was collected in 2007-2008. However, in some instances where 2007-2008 data were not available, the most recent information was included.

### *Canada*

Geographically, Canada is located in North America with a total area of 9,984,670 square kilometers. The estimated population for Canada in 2008 was

33,212,696. Canadian people consisted of several ethnic groups such as British Isles (28%), French (23%), other European (15%), Amerindian (2%), Other, mostly Asian, African, and Arab (6%), and mixed background (26%). There are two official languages in Canada with English spoken by 59.3% and French by 23.2% of the people. The remaining 17.5% speak other languages (Central Intelligence Agency, 2008).

Canada's political system is a Constitutional monarchy that is also a parliamentary democracy and a federation. The Chief of state is the Queen who is represented by the Governor and the Head of government who is known as the Prime Minister. Canada is an affluent, high-tech industrial society in the trillion-dollar class with a market-oriented economic system. In 2007, GDP for Canada was $38,200 (Central Intelligence Agency, 2008).

Education in Canada is provided, funded and overseen by federal, provincial, and local government. Education is compulsory up to the age of 16 in every province, except for Ontario and New Brunswick, where the compulsory age is 18. In general, the educational system in Canada is divided into Elementary (Primary School), followed by Secondary (High School) and Post-Secondary (University and College) (Education in Canada, 2008).

In recent years, the Canadian education system focused its attentions on several issues and problems such as deprofessionalization; the dominance of a political-economic imperative in the formulation of state educational policy (accountability, privatization, market, choice, and decentralization); multiculturalism and diversity; restructuring and retrenchment; and the demographic changes facing all industrialized nations (Education in Canada, 2008).

*The United States*

Geographically, the United States is located in North America with a total area of 9,826,630 square kilometers. The estimated population for the United States in 2008 was 303,824,646. The United States is made up of several ethnic groups such as Caucasian (81.7%), African American (12.9%), Asian (4.2%), Amerindian and Alaska native (1%), native Hawaiian and other Pacific islander (0.2%). The official language in the United States is English, which is spoken by more than 82% of the people. Other spoken languages in the United States include Spanish (10.7%), other Indo-European (3.8%), Asian and Pacific island (2.7%), and Other (0.7%) (Central Intelligence Agency, 2008).

The United States' political system is a Constitution-based federal republic with a strong democratic tradition. The Chief of state is the President and the Head of government is also the President. In the United States, the President appoints the Cabinet. The United States is known as the largest and most technologically powerful economy in the world with a market-oriented economy system. In 2007, GDP for the United States was $46,000 (Central Intelligence Agency, 2008).

Education in the United States is generally divided into Elementary (Primary School), followed by Junior and High School and Post-Secondary (University and College). School attendance is mandatory at the elementary, junior and high school levels. The ages for compulsory education vary by state, beginning at age of five to eight and ending at the age of fourteen to eighteen. Students are placed in year groups known as grades, beginning with first grade and culminating in twelfth grade. A growing number of states are now requiring school attendance until the age of 18 (Education in the United States, 2008).

Education in the United States is provided mainly by the government, with control and funding coming from three levels: federal, state, and local. In 2005, the United States ranked the first in the world in terms of annual spending per student on its public schools (approximately $11,000 per student) (Education in the United States, 2008).

In recent years, in the United States, major educational issues centered on curriculum, funding, and control. Of critical importance, because of its enormous implications on education and funding, is the No Child Left Behind Act of 2002. Under this Act, schools are held accountable for meeting the learning standards that are set by the state and school districts in the areas of reading, writing, math, and science (Education in the United States, 2008).

*Egypt*

Geographically, Egypt is located in North Africa with a total area of 1,001,450 square kilometers. The estimated population for Egypt in 2008 was 81,713,517. In Egypt, Egyptians accounted for 98% of the population. The remaining consisted of Berber, Nubian, Bedouin, and Beja (1%) and Greek, Armenian, other European (primarily Italian and French) (1%). The official language in Egypt is Arabic. However, English and French are also widely understood by educated classes (Central Intelligence Agency, 2008).

Egypt is a republic country. The Chief of state is the President and the Head of government is the Prime Minister. In Egypt, the Cabinet is appointed by the President. According to the Central Intelligence Agency (2008), Egypt's economy depends mainly on agriculture, media, petroleum exports, and tourism. Recently, the government has struggled to prepare the economy for the new millennium through economic reform and massive investments in communications and physical infrastructure. In 2007, GDP for Egypt was $5,400 (Central Intelligence Agency, 2008)

Education in Egypt is highly centralized, and is divided into three stages: Basic Education, Secondary Education, and Post-Secondary education. Basic education includes six years of primary school and 3 years of intermediate school. Promotion from primary to intermediate school is determined by examination scores. Since 1981, the government in Egypt issued a law that stated that basic education is free and compulsory to all students ages 6 through 14. Beyond this stage, education depends on the student's ability (Education in Egypt, 2008).

In Egypt, schools are referred to as government schools or private schools. There are two types of government schools: (1) the Arabic schools which provide the governmental national curriculum in the Arabic language and (2) the experimental language schools which teach most of the government curriculum in English, and add French as a second foreign language. As for private schools, there are three types: (1) ordinary schools which are quite similar to that of the government schools, but pay more attention to the students' personal needs and to the school facilities; (2) the language schools teach most of the government curriculum in English, and add French or German as a second foreign language; (3) the religious schools are religiously oriented schools and their curricula differ from the remaining schools. In Egypt, the enrollment rate for girls is significantly lower than for boys. Overall students’ attendance rate was also low (Education in Egypt, 2008).

*South Africa*

Geographically, South Africa is located in South Africa with a total area of 1,219,912 square kilometers. The estimated population for South Africa in 2008 was 43,786,115. In South Africa, black African accounted for 79%, White 9.6%, colored 8.9%, and Indian/Asian 2.5%. There are many languages currently spoken in South Africa: IsiZulu by 23.8% of the people, IsiXhosa 17.6%, Afrikaans 13.3%, Sepedi 9.4%, English 8.2%, Setswana 8.2%, Sesotho 7.9%, Xitsonga 4.4%, and other 7.2% (Central Intelligence Agency, 2008).

South Africa is a republic country. The Chief of state is the President and the Head of government is also the President. In South Africa, the Cabinet is appointed by the President. According to the Central Intelligence Agency (2008), South Africa has a market economy with an abundant supply of natural resources, well-developed financial, legal, communications, energy, and transport sectors. However, economic problems remain from the apartheid era (e.g., poverty and lack of economic empowerment among the disadvantaged groups). In 2007, GDP for South Africa was $10,600 (Central Intelligence Agency, 2008).

Education in South Africa has three levels: General Education and Training from grade 0 through grade 9, Further Education and Training from grade 10 through grade 12, and Higher Education and Training for technical schools and college and university. Under the South African Schools Act of 1996, education is compulsory for all South Africans from age 7 (grade 1) to age 15, or grade 9. In South Africa, the government provides a national framework for school policy and spends approximately 20% of their expenditure on education annually. Students can choose to attend public schools which are funded by the government or private schools where they have to pay for education (Education in South Africa, 2006).

The major educational issues faced by South Africa in recent years include imbalances in education remaining from the apartheid legacy, increasing dropout rates for girls, and discrepancy in educational opportunities between rural and urban schools. Recent statistics suggested that among the South African population, only 14% of black Africans have an education of high school or higher, whereas 40% of Indians and 65% of Whites have an education of high school or higher (Education in South Africa, 2006).

## *Instruments*

### *Eighth-grade Mathematics Assessment Survey*

TIMSS 2003 eighth-grade math assessment was a successor of TIMSS 1995 and 1999 eighth-grade math assessments, and thus the curriculum framework and test booklet design used in 1995 and 1999 were also used in 2003. This was to ensure reliable measurement of trends in math teaching and learning over time. However, because a large number of the items on the TIMSS 1995 and 1999 were released for public use after each cycle of the assessment, new items were developed to replace the retired items in the TIMSS 2003 eighth-grade math assessment (Martin, 2005). According to Martin (2005), of the 426 score points available in the entire 2003 eight grade mathematics and science assessment, 47 came from items used also in 1995, 102 from items used also in 1999, and 267 from items used for the first time in 2003.

*Test booklet.* The TIMSS assessment framework employed a matrix-sampling technique that assigned each assessment item to one of a set of item blocks, and then assembled student test booklets by combining the item blocks according to a balanced design in order to achieve broad subject matter coverage (Martin, 2005). For the TIMSS 2003 eighth-grade math assessment, a total of 194 items were categorized into 14 blocks which were labeled M1 through M14. Blocks 1 through 6 contained secure items from earlier TIMSS assessments (TIMSS 1995 and 1999) to measure trends and blocks 7 through 14 contained new replacement items. These 14 blocks of items then were distributed across 12 student booklets. Each booklet consisted of two to four blocks of items, resulting in a different number of items in each booklet. Each student was randomly assigned one booklet. A summary of the TIMSS 2003 eighth grade mathematics assessment booklet matrix is presented in Table 2.

Table 2.
*TIMSS 2003 Eighth-Grade Math Assessment Booklet Assembling Matrix*

| | Block | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Booklet | M 1 | M 2 | M 3 | M 4 | M 5 | M 6 | M 7 | M 8 | M 9 | M 10 | M 11 | M 12 | M 13 | M 14 |
| 1 | ■ | ■ | | | ■ | | ■ | | | | | | | |
| 2 | | ■ | ■ | | | ■ | | ■ | | | | | | |
| 3 | | | ■ | ■ | | | | | | | ■ | | ■ | |
| 4 | | | | ■ | ■ | | | | | | | ■ | | ■ |
| 5 | | | | | ■ | ■ | | | ■ | | | | ■ | |
| 6 | ■ | | | | | ■ | | | | ■ | | | | ■ |
| 7 | | | | | | ■ | ■ | | | | | | | |
| 8 | | | | | ■ | | | ■ | | | | | | |
| 9 | | | | ■ | | | | | ■ | | | | | |
| 10 | | | ■ | | | | | | | ■ | | | | |
| 11 | | ■ | | | | | | | | | ■ | | | |
| 12 | ■ | | | | | | | | | | | ■ | | |

As a result of the matrix-sampling design, the number of assessment items by booklet and domain varied considerably, ranging from 26 to 60 items per booklet. Table 3 displays a summary of TIMSS 2003 eighth grade math item breakdowns by assessment domain and booklet.

Table 3.
*Number of Items by Domain and Booklet in TIMSS 2003 Eighth-Grade Math Assessment*

| | Booklet | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Domain | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Algebra | 11 | 12 | 14 | 17 | 7 | 12 | 3 | 6 | 8 | 6 | 7 | 8 |
| Data | 9 | 8 | 5 | 8 | 4 | 9 | 6 | 3 | 2 | 5 | 4 | 6 |
| Geometry | 8 | 11 | 9 | 6 | 12 | 9 | 5 | 4 | 5 | 4 | 5 | 5 |
| Measurement | 10 | 7 | 11 | 11 | 9 | 5 | 3 | 6 | 8 | 4 | 5 | 4 |
| Number | 17 | 22 | 18 | 13 | 23 | 17 | 10 | 9 | 8 | 7 | 8 | 7 |
| Total Items | 55 | 60 | 57 | 55 | 55 | 52 | 27 | 28 | 31 | 26 | 29 | 30 |

*Subject content areas.* The TIMSS 2003 eighth-grade math assessment contained five content areas: number, algebra, measurement, geometry, and data. The number content domain consisted of understandings and skills related to whole numbers, fractions and decimals, integers, ratio, proportion, and percent. The major topic areas in algebra were patterns, algebraic expressions, equations and formulas, and relationships. The measurement content domain included attributes and units, tools, techniques, and formulas. For geometry, five major topics were included: lines and angles, two- and three-dimensional shapes, congruence and similarity, locations and spatial relationships, and symmetry and transformations. Finally, the data content domain consisted of four topic areas: data collection and organization, data representation, data interpretation, and uncertainty and probability (Martin, 2005).

*Item writing and development.* The development of the TIMSS 2003 eighth-grade math assessment was a collaborative process spanning more than two years, from September 2000 to March 2003, and involving math educators and development specialists from all participating countries (TIMSS, 2003). The eighth-grade math assessment development was guided by the assessment framework and specifications which focused on two dimensions: content domains and cognitive domains. There were five content domains: number, algebra, measurement, geometry, and data. There were four cognitive domains: knowing facts and procedures, using concepts, solving routine problems, and reasoning.

With support and training from the TIMSS International Study Center, National Research Coordinators (NRCs) from participating countries contributed a large pool of items for review and field testing. The International Study Center established a math task force to manage the item development process. To help review, select and revise items for the assessment and to ensure their mathematical accuracy, the International Study Center convened the Math Item Review Committee, an international committee of prominent mathematics experts nominated by participating countries and representing a range of nations and cultures. As a result of this item development process, more than 2000 draft items covering a wide array of topics and a range of cognitive domains and item types were submitted to the item pool for further review by the Math Item Review Committee. Because the items were developed in English and translated into 34 languages by the participating countries, both the Math Item Review Committee and National Research Coordinators were important in identifying any items that might prove difficult to translate consistently (TIMSS, 2003).

Of the new items developed, 190 were selected for the field test in 41 countries. International item analysis of the results from the field test was used to inform the review and selection of items for the main survey. For the final TIMSS 2003 eighth-grade assessment, there were a total of 194 items (115 newly developed items and 79 trend items). Of these items, 128 were multiple choice items and 66 were constructed-response items (TIMSS, 2003).

*Item types.* Two item types were used in the TIMSS 2003 eighth-grade math assessment: multiple choice and constructed-response. For constructed-response items, students were required to construct a written response, rather than select a response from a set of options like in multiple-choice items. Correct answers for multiple-choice item were credited one point. Constructed-response items were awarded one, two, or three points, depending on the nature of the task and the skills required to complete it. Up to two-thirds of the total number of points represented by all the items came from multiple-choice items (Martin, 2005).

*Translation, cultural adaptation, and verification.* The TIMSS 2003 data collection instruments (achievement tests and background questionnaires) were originally developed and prepared in English and subsequently translated by the participating countries into 34 national languages of instruction and cultural contexts. To control the quality of translated versions, each translation went through a rigorous verification process that included verification by an international translation company, review by the International Study Center, verification of the item translations at the national centers and a check by International Quality Control Monitors. The goal of this translation, adaptation, and verification process was to ensure that translated instruments were accurately and internationally comparable (Martin et al., 2004).

*Reliability estimates.* Item analysis and review were conducted internally for the TIMSS 2003 achievement data in order to examine and evaluate the psychometric characteristics of each achievement item in all participating countries. For all items, regardless of item format, multiple statistics were computed to yield information about the reliability estimates of an item. These statistics include: (1) the number of students that responded in each country, (2) the difficulty level (the percentage of students that answered the item correctly), (3) the Rasch one-parameter IRT item difficulty index, (4) the discrimination index (the point-biserial correlation between success on the item and a total score), (5) the distracter index (the percentage of students that selected each of the distracters), (6) the reliability-score (the percentage of exact agreement between two independent scorers), (7) the item-by-country interaction index (i.e., when a high-scoring country has low performance on an item on which other countries are doing well, there is said to be an item-by-country interaction), (8) the scoring reliability (i.e., a particular student response should receive the same score, regardless of scorer), (9) the within-country scoring reliability (i.e., a random sample of at least 200 student responses to each item per country was selected to be scored independently by two scorers), (10) trend item scoring reliability (i.e., the percentage of exact agreement between scorers across years for the same item), and (11) the cross-country scoring reliability (the percentage of exact agreement among scorers in 20 English-speaking countries) (TIMSS, 2003).

Of these statistics, the international means of the item difficulties and item discriminations served as guides to the overall statistical properties of the items. For TIMSS 2003 eighth-grade math assessment, the international mean of the item difficulties was .66,  mean Rasch difficulties was -1.46, and mean item discriminations was .44, which indicates appropriately reliable assessment items. In addition, the international average of exact percent agreement across items was high, 99%. Similarly, the scorer reliability across English-speaking countries was high, with the percent exact agreement averaging 96% across the 20 math items (TIMSS, 2003).

*Reported achievement scores.* TIMSS 2003 reported trends in student achievement in both the general area of math and in the major subject matter content areas. Because each student responded to only part of the assessment, these parts had to be combined for an overall picture of the assessment results for each country. Using item response theory (IRT) methods, individual student responses to math items were placed on common scales that link to TIMSS results from 1995 and 1999 to track their progress in math achievement since then. A three-parameter IRT model was applied to multiple-choice items which were dichotomously scored (correct or incorrect). For constructed-response items with 0, 1, or 2 available score points, a generalized partial credit models was used. The IRT scaling method produced a score by averaging the responses of each student to the items that he or she took in a way that takes into account the difficulty and discriminating power of each item (Martin, 2005).

*Raw scores.* For TIMSS 2003 eighth-grade math assessment, raw scores were computed by adding the number of points obtained by each student over all the items in the student's test booklet. Multiple-choice items were scored 1 for correct answers and 0 for incorrect answers and constructed-response items were scored 0 for incorrect answers, 1 for partially correct answers, and 2 for correct answers. Because the raw score was dependent on the number of items in the student's test book, and because the number of items varied from test book to test book, the raw scores can be used only to compare students' performance on the same booklet in the same year (Martin, 2005). Table 4 shows the maximum number of score points for eighth-grade math by booklet and by domain content.

Table 4.
*Maximum Number of Score Points in TIMSS 2003 Eighth-Grade Math Assessment*

| | Booklet | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Domain | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Number | 19 | 22 | 19 | 14 | 25 | 17 | 11 | 10 | 8 | 7 | 8 | 7 |
| Algebra | 12 | 12 | 15 | 22 | 7 | 15 | 4 | 6 | 9 | 6 | 7 | 9 |
| Measurement | 13 | 7 | 11 | 13 | 11 | 5 | 4 | 8 | 8 | 4 | 5 | 4 |
| Geometry | 8 | 11 | 11 | 7 | 14 | 10 | 5 | 4 | 5 | 4 | 5 | 5 |
| Data | 11 | 8 | 6 | 8 | 4 | 12 | 8 | 3 | 2 | 8 | 5 | 6 |
| Total scores | 63 | 60 | 62 | 64 | 61 | 59 | 32 | 31 | 32 | 29 | 30 | 31 |

*Standardized raw scores.* In order to improve the utility of students' achievement scores, raw scores were standardized by booklet to provide a simple score that could be used in comparisons across booklets in the same year. The standardized score had the weighted mean score of 50 and a weighted standard deviation of 10 within each booklet in a country (Martin, 2005).

*National Rasch scores.* Based on the one-parameter Rasch model with maximum likelihood (ML) estimation, the national Rasch scores were standardized to have a mean score of 150 and a standard deviation of 10 within each country. The main purpose of national Rasch scores was to provide a preliminary measure of overall math achievement that could be used as a criterion variable in studies of item discrimination prior to the TIMSS 2003 IRT scaling. Because each country has the same mean score and dispersion, these scores should not be used for international comparison (Martin, 2005).

*Plausible values.* Due to the TIMSS 2003 assessment design, each student only responded to the items on one test booklet or a subset of the item pool. In order to derive estimates for each student of the overall score they would have achieved had they completed the entire assessment, TIMSS 2003 used a sophisticated psychometric scaling technique (known as item response theory scaling with conditioning and multiple imputation) to generate imputed scores for those items that were not administered to the student. Because there was some error inherent in the imputation process, the TIMSS database provided five separate plausible values for each student on each of the scales. In other words, each student had five plausible values of his or her achievement on the overall math and five plausible values on each content domain area. Overall plausible values were standardized with a mean of 500 and standard deviation of 100 and may be compared across test administrations. Content domain plausible values, however, should not be compared across test administration because when standardized they have different means and standard deviations (Martin, 2005).

*Background Surveys*

The background questionnaires are based on the TIMSS 2003 Contextual Framework, which specifies the major characteristics of the educational and social contexts to be studied and identifies the areas to be addressed in the background questionnaires. The background questionnaires were developed by an expert committee composed of international educators and measurement specialists with much input from the National Research Coordinators (Martin, 2005). The administration of the student questionnaires was conducted at participating schools by test administrators who also administered the student test booklets. As for the teacher questionnaires and school questionnaires, the school coordinators distributed these background questionnaires to corresponding teachers and school principals and made sure that the questionnaires were returned completed.

*Eighth-grade mathematics student background survey.* The student mathematics questionnaire had a total of 18 forced choice questions that sought information about the students' demographic background, home resources, their experiences in learning mathematics, and their perceptions about school environment. It is worthy of note, however, that due to cultural differences, some questions or some question options in the student questionnaire were adapted or even removed to fit with the national contexts. For example, for the question “Do you have any of these items at home?”, there were a total of 16 options. Of these options, four were fixed (i.e., calculator, computer, study desk, and dictionary) and the remaining options were adapted to include common country-specific items. Some countries (e.g., Australia, Bulgaria, Chile) might also opt to include fewer than 16 options in this question.

*Eighth-grade mathematics teacher survey.* The teacher questionnaire included 28 forced choice questions in order to gather information about the teachers’ preparation and professional development, their pedagogical activities, and the implemented curriculum. Like in the student background questionnaire, due to cultural differences, some questions or some question options in this questionnaire were adapted or even removed to fit with the national contexts. For example, the question “What requirements did you have to satisfy in order to become a mathematics teacher at grade 8?” was not administered in the U.K and thus, data for this variable were not available for the U.K.

*School survey.* The school questionnaire contained 18 questions asking school principals or headmasters to provide information about the school contexts for the teaching and learning. Similar to the previous background questionnaires, not all of the questions in the school questionnaire were administered in all of the countries that participated in the TIMSS 2003, resulting in lack of data for some school-related variables in some countries.

*Variables*

The selection of variables for this study was guided by the conceptual model (Carroll, 1963), the review of existing literature on contextual factors related to student math achievement (see Chapter Two), and the practical implications of the variables to policy issues. The dependent variable of the study is Overall Mathematics Score, an IRT-based score, which was calculated by averaging five plausible sub-topic scores: algebra, number, geometry, measurement, and data.

The independent variables are five groups of factors: (1) student background, (2) home resources, (3) instructional practices, (4) teacher background, and (5) school background. Each of these groups of factors was precisely defined by using existing variables in the TIMSS 2003 database. For example, student background was measured by five variables: gender, self confidence in math, valuing of math, time on math homework, and extra math lessons. Home resources was represented by three variables indicating the availability of: calculator, computer, and desk for student's use at home. Instructional practices had nine indicators: opportunity to learn number, opportunity to learn algebra, opportunity to learn measurement, opportunity to learn geometry, opportunity to learn data, amount of homework assignment, content-related activities in math lessons, instructional practice-related activities in math lessons, and instructional time. Teacher background was represented by preparation to teach, ready to teach number, ready to teach algebra, ready to teach measurement, ready to teach geometry, ready to teach data, and math-related professional development. Finally, school background was measured by class size, school resources for math instruction, and teacher's perception of math instructional limitations due to student factors.

Table 5 presents a mapping between the variables selected for this study and the variables identified in Carroll's (1963) conceptual model of school learning. It is worthy of note that because the variables used in this study were selected from existing secondary database as opposed to being created for primary research, some loose connections between these variables and those from the model were anticipated.

Table 5.
*Mapping of Variables in Caroll's Model With Variables in the Study*

| Variables in Caroll's Model | Variables in the study |
|---|---|
| 1) Aptitude – the amount of time needed to learn the task under optimal instructional conditions<br><br>2) Ability to understand instruction<br><br>3) Perseverance – the amount of time the learner is willing to engage actively in learning | 1) Student background (self-confidence in learning math, valuing math, time on math homework, tutoring in math, and gender)<br>2) Home resources (availability of calculator, dictionary, computer, and desk at home for student's use) |
| 4) Opportunity to learn – time allowed for learning | 3) Instructional practices (opportunity to learn in terms of topic coverage before the time of the test, amount of homework assignment, activities in math lessons, and average of math instructional hours per year) |
| 5) Quality of instruction – the extent to which instruction is presented so that no additional time is required for mastery beyond that required in regard to aptitude | 4) Teacher background (preparation to teach, ready to teach, and professional development)<br>5) School background (class size for math instruction, school resources for math instruction, and teachers' perceptions of math instructional limitations due to student factors) |

Table 6 presents an explicit description of the study's contextual and background variables and their respective indicators. In this table, a composite variable was marked as "TIMSS derived variable" if it was provided in the TIMSS database and as "computed by researcher" if it was created by the researcher. One exception occurred when the researcher renamed the TIMSS derived variable from "Math classes with few or no instructional limitation due to student factor" with " teacher's perception of math instructional limitations due to student factors" to improve the meaningfulness of the variable name.

Table 6.
*Description of Contextual and Background Variables*

| Variable Name | Variable Description |
|---|---|
| | Student Background |
| Gender of student | Are you a girl or a boy?<br>1 = girl, 2 = boy |
| Student self-confidence in learning math<br><br>(TIMSS derived variable) | Composite variable ranging from 1-3 (high, medium, low). Four items were used to create the composite variable.<br>How much do you agree with these statements about learning mathematics? (4-point scale: agree a lot, agree a little, disagree a little, and disagree a lot).<br>1) I usually do well in math<br>2) Math is more difficult for me than for many of my classmates<br>3) Math is not one of my strengths<br>4) I learn things quickly in math |
| Student valuing math<br><br>(TIMSS derived variable) | Composite variable ranging from 1-3 (high, medium, low). Seven items were used to create the composite variable.<br>How much do you agree with these statements about learning mathematics? (4-point scale: agree a lot, agree a little, disagree a little, and disagree a lot).<br>1. I would like to take more math in school<br>2. I enjoy learning math<br>3. I think learning math will help me in my daily life<br>4. I need math to learn other school subjects<br>5. I need to do well in math to get into the university of my choice<br>6. I would like a job that involved using math<br>7. I need to do well in math to get the job I want |
| Time on math homework | On a normal school day, how much time do you spend before or after school doing mathematics homework? (5-point scale: 1 = no time, 2 = less than one hour, 3 = 1-2 hours, 4 = more than 2 but less than 4 hours, and 5 = 4 or more hours) |
| Tutoring/Extra math lessons | During this school year, how often have you had extra lessons or tutoring in mathematics that is not part of your regular class? (4-point scale: 1 = every or almost every day, 2 = once or twice a week, 3 = sometimes, 4 = never or almost never) |
| | Home Resources |
| Home resources for learning<br><br>(computed by researcher) | Composite variable, a sum of students' responses for three variables.<br>Do you have any of these items at home? (2-point scale 1 = yes, 2 = no )<br>Calculator<br>Computer (excluding Xbox, playstation or TV/Video game computer)<br>Desk |
| | Instructional Practices |
| Opportunity to | Composite variable, an average percent of students whose teachers checked option 1 |

Table 6.
*Description of Contextual and Background Variables*

| Variable Name | Variable Description |
|---|---|
| learn number<br><br>(TIMSS derived variable) | (mostly taught before this year) and option 2 (mostly taught this year) for the 10 items of the number domain. Details of these 10 items can be found in Appendix B.<br>The following list includes the main topics addressed by the TIMSS mathematics test. Choose the response that best describes when students in the TIMSS class have been taught each topic. If a topic was taught half this year and half before this year, please choose "Mostly taught this year." (3 point-scale: 1 = mostly taught before this year, 2 = mostly taught this year, and 3 = not yet taught or just introduced) |
| Opportunity to learn algebra<br><br>(TIMSS derived variable) | Composite variable, an average percent of students whose teachers checked option 1 (mostly taught before this year) and option 2 (mostly taught this year) for the 6 items of the algebra domain. Details of these 6 items can be found in Appendix B.<br>The following list includes the main topics addressed by the TIMSS mathematics test. Choose the response that best describes when students in the TIMSS class have been taught each topic. If a topic was taught half this year and half before this year, please choose "Mostly taught this year." (3 point-scale: 1 = mostly taught before this year, 2 = mostly taught this year, and 3 = not yet taught or just introduced) |
| Opportunity to learn measurement<br><br>(TIMSS derived variable) | Composite variable, an average percent of students whose teachers checked option 1 (mostly taught before this year) and option 2 (mostly taught this year) for the 8 items of the measurement domain. Details of these 8 items can be found in Appendix B.<br>The following list includes the main topics addressed by the TIMSS mathematics test. Choose the response that best describes when students in the TIMSS class have been taught each topic. If a topic was taught half this year and half before this year, please choose "Mostly taught this year." (3 point-scale: 1 = mostly taught before this year, 2 = mostly taught this year, and 3 = not yet taught or just introduced) |
| Opportunity to learn geometry<br><br>(TIMSS derived variable) | Composite variable, an average percent of students whose teachers checked option 1 (mostly taught before this year) and option 2 (mostly taught this year) for the 13 items of the geometry domain. Details of these 13 items can be found in Appendix B.<br>The following list includes the main topics addressed by the TIMSS mathematics test. Choose the response that best describes when students in the TIMSS class have been taught each topic. If a topic was taught half this year and half before this year, please choose "Mostly taught this year." (3 point-scale: 1 = mostly taught before this year, 2 = mostly taught this year, and 3 = not yet taught or just introduced) |
| Opportunity to learn data<br><br>(TIMSS derived variable) | Composite variable, an average percent of students whose teachers checked option 1 (mostly taught before this year) and option 2 (mostly taught this year) for the 8 items of the data domain. Details of these 8 items can be found in Appendix B.<br>The following list includes the main topics addressed by the TIMSS mathematics test. Choose the response that best describes when students in the TIMSS class have been taught each topic. If a topic was taught half this year and half before this year, please choose "Mostly taught this year." (3 point-scale: 1 = mostly taught before this year, 2 = mostly taught this year, and 3 = not yet taught or just introduced) |
| Amount of homework assignment | Composite variable with 3 point-scale: 1 = high, 2 = medium, 3 = low. This composite variable was created using three variables.<br>1) Do you assign mathematics homework to the TIMSS class? (2 point-scale: 1 = yes, 2 = no) |

Table 6.
*Description of Contextual and Background Variables*

| Variable Name | Variable Description |
|---|---|
| (TIMSS derived variable) | 2) How often do you usually assign mathematics homework to the TIMSS class? (3 point-scale: 1 = Every or almost every lesson, 2 = About half the lessons, 3 = some lessons<br>3) When you assign mathematics homework to the TIMSS class, about how many minutes do you usually assign? (Consider the time it would take an average student in your class.) (5-point scale: 1 = Fewer than 15 minutes, 2 = 15-30 minutes, 3 = 31-60 minutes, 4 = 61-90 minutes, 5 = More than 90 minutes) |
| Content-related activities in math lessons<br><br>(computed by researcher) | Composite variable computed by averaging student responses for the following 4 items.<br>How often do you do these things in your mathematics lesson? (4-point scale: 1 = every or almost every lesson , 2 = about half the lessons, 3 = some lessons, and 4 = never)<br>1. We practice adding, subtracting, multiplying, and dividing without using a calculator<br>2. We work on fractions and decimals<br>3. We interpret data in tables, charts, or graphs,<br>4. We write equations and functions to represent relationships |
| Instructional practice-related activities in math lessons<br><br>(computed by researcher) | Composite variable computed by averaging student responses for the following 9 items.<br>How often do you do these things in your mathematics lesson? (4-point scale: 1 = every or almost every lesson , 2 = about half the lessons, 3 = some lessons, and 4 = never)<br>1. We work together in small groups<br>2. We relate what we are learning in mathematics to our daily life<br>3. We explain our answers<br>4. We decide on our own procedures for solving complex problems<br>5. We review our homework<br>6. We listen to the teacher give a lecture-style presentation<br>7. We work problems on our own<br>8. We begin our homework in class<br>9. We have a quiz or test |
| Average math instructional hours per year<br><br>(TIMSS derived variable) | Composite variable computed from the following variables.<br>From the school survey:<br>1) How many days per year is your school open for instruction for eighth-grade students?<br>2) How many instructional days are there in the school week (typical calendar week from Monday through Saturday) for eighth-grade students? (Seven-point scale from none – 6 days)<br>2A) Number of full days (over 4 hours)?<br>2B) Number of half days (4 hours or less)?<br>3) To the nearest half-hour, what is the total instructional time in a typical full day (excluding lunch break, study hall, and after school activities) for eighth-grade students? (Six-point scale: 1 = 4 hours or less, 2 = 4.5 hours, 3 = 5 hours, 4 = 5.5 hours, 5 = 6 hours, 6 = 6.5 hours or more)<br>4) How many full instructional days?<br>From the teacher survey:<br>1) How many minutes per week do you teach math to the TIMSS class? |
| | Teacher Background |
| Preparation to | During your post-secondary education, what was your major or main area(s) of |

Table 6.
*Description of Contextual and Background Variables*

| Variable Name | Variable Description |
|---|---|
| teach math content (computed by researcher) | study?<br>1) Mathematics: 1 = yes, 2 = no<br>2) Education – Mathematics: 1 = yes, 2 = no |
| Ready to teach number<br><br>(computed by researcher) | Composite variable computed by averaging math teacher responses for 2 items of the number domain. Details of these 2 items can be found in Appendix C.<br>Considering your training and experience in both mathematics content and instruction, how ready do you feel you are to teach each topic at eighth grade? ( 3-point scale: 1 = very ready, 2 = ready, 3 = not ready). |
| Ready to teach algebra<br><br>(computed by researcher) | Composite variable computed by averaging math teacher responses for 4 items of the algebra domain. Details of these 4 items can be found in Appendix C.<br>Considering your training and experience in both mathematics content and instruction, how ready do you feel you are to teach each topic at eighth grade? ( 3-point scale: 1 = very ready, 2 = ready, 3 = not ready). |
| Ready to teach measurement topic<br><br>(computed by researcher) | Composite variable computed by averaging math teacher responses for 4 items of the measurement domain. Details of these 4 items can be found in Appendix C.<br>Considering your training and experience in both mathematics content and instruction, how ready do you feel you are to teach each topic at eighth grade? ( 3-point scale: 1 = very ready, 2 = ready, 3 = not ready). |
| Ready to teach geometry<br><br>(computed by researcher) | Composite variable computed by averaging math teacher responses for 4 items of the geometry domain. Details of these 4 items can be found in Appendix C.<br>Considering your training and experience in both mathematics content and instruction, how ready do you feel you are to teach each topic at eighth grade? ( 3-point scale: 1 = very ready, 2 = ready, 3 = not ready). |
| Ready to teach data<br><br>(computed by researcher) | Composite variable computed by averaging math teacher responses for 4 items of the data domain. Details of these 4 items can be found in Appendix C.<br>Considering your training and experience in both mathematics content and instruction, how ready do you feel you are to teach each topic at eighth grade? ( 3-point scale: 1 = very ready, 2 = ready, 3 = not ready). |
| Math-related professional development<br><br>(computed by researcher) | In the past two years, have you participated in professional development in any of the following?<br>1) Math content: 1 = yes, 2 = no<br>2) Math pedagogy/instruction: 1 = yes, 2 = no<br>3) Math curriculum: 1 = yes, 2 = no<br>4) Math assessment: 1 = yes, 2 = no<br>5) Problem solving/critical thinking: 1 = yes, 2 = no |
| | School Background |
| Class size for math instruction<br><br>(TIMSS derived variable) | How many students are in the TIMSS class?<br>Four categories were derived based on the teachers' responses<br>1 = 1-24 students<br>2 = 25-32 students<br>3 = 33-40 students<br>4 = 41 or more students |
| School resources | Composite variable with 3 point-scale: 1 = high, 2 = medium, 3 = low. This |

Table 6.
*Description of Contextual and Background Variables*

| Variable Name | Variable Description |
|---|---|
| for math instruction<br><br>(TIMSS derived variable) | composite variable was computed from the school principals' responses to the 10 following items:<br>Is your school's capacity to provide instruction affected by a shortage or inadequacy of any of the following? (4-point scale: 1 = none, 2 = a little, 3 = some, 4 = a lot).<br>1) Instructional materials (e.g., textbook)<br>2) Budget for supplies (e.g., paper, pencils)<br>3) School buildings and grounds<br>4) Heating/cooling and lighting systems<br>5) Instructional space (e.g., classrooms)<br>6) Computers for mathematics instruction<br>7) Computer software for mathematics instruction<br>8) Calculators for mathematics instruction<br>9 ) Library materials relevant to mathematics instruction<br>10) Audio-visual resources for mathematics instruction |
| Teacher' perception of math instructional limitations due to student factors<br><br>(TIMSS derived variable) | Composite variable with 3 point-scale: 1 = high, 2 = medium, 3 = low. This composite variable was computed by averaging teachers' responses to six following items.<br>In your view, to what extent do the following limit how you teach the TIMSS class? (5-point scale: 1 = not applicable , 2 = not at all, 3 = a little, 4 = some, 5 = a lot)<br>1) Students with different academic abilities<br>2) Students who come from a wide range of backgrounds (e.g., economic, language)<br>3) Students with special needs, (e.g., hearing, vision, speech impairment, physical disabilities, mental or emotional/psychological impairment)<br>4) Uninterested students<br>5) Low morale among students<br>6) Disruptive students |

The variables instructional practices, teacher background and school background in this study can be manipulated by state, district and school policies. It is possible to create more effective mathematics lessons, to increase opportunities to learn, to better utilize school resources, and to adjust class size and mathematics instructional practices so as to stimulate teaching and learning mathematics in a school. Although the predictor variables student background and home resources appear to be more difficult to manipulate by policy, they provide educators and parents with important information about student-related factors that can influence mathematics achievement.

*Reliability of Composite Predictor Variables*

Of the contextual variables listed in Table 6, 18 were composite variables created through principal factor analysis with promax rotation. Specifically, self-confidence was created from four variables related to students' reported level of confidence in learning math. Through an examination of the scree plot, eigenvalues, and interpretability of variables, a single factor was retained. All the items with factor pattern coefficient larger than .30 were included in the computation of the composite variable which had Cronbach's $\alpha$ of .73. Appendix D provides details of item factor pattern coefficients.

The composite variable, valuing of math, was constructed from seven items related to how students perceive the importance of learning math. One single factor was retained as the result of examining the scree plot, eigenvalues, factor pattern coefficients greater than or equal to .30, and interpretability of items. This composite variable had a Cronbach's $\alpha$ of .79. Details of item factor pattern coefficients can be found in Appendix D.

The next composite variable is home resources for learning which was measured by three items related to availability of study aids (i.e., calculator, computer, and desk) for students to use at home. An evaluation of the scree plot, eigenvalues, factor pattern coefficients greater than .30 and interpretability of items suggested that these items were unidimentional. However, Cronbach's $\alpha$ for this composite variable was relatively modest, .44. Details of item factor pattern coefficients can be found in Appendix D.

Five composite variables (i.e., Opportunity to learn algebra, number, geometry, measurement, and data) were constructed from a series of 45 items related to teachers' responses as to when during the school year each of the math topics was taught to the

TIMSS class. Specifically, opportunity to learn number was measured by 10 items, opportunity to learn algebra by six items, opportunity to learn measurement by eight items, opportunity to learn geometry by 12, and opportunity to learn data by eight items. Using similar factor analysis criteria and setting the number of factor equal to 1, these five composite variables were essentially unidimentional. All the items measuring each of the composite variables had factor pattern coefficients larger than .30 and the obtained Cronbach's $\alpha$ for these composite variables were relatively high, ranging from .74 to .91. Details of item factor pattern coefficients can be found in Appendix D.

Two composite variables were created from nine items related to students' reported activities in math lessons. Instructional practice-related activities in math lessons were measured by 5 items and content-related activities in math lessons were measured by four items. Similar criteria were used to determine item inclusion in the factor (i.e., examination of the scree plot, eigenvalues, factor pattern coefficients greater than or equal to .30, and interpretability of items). As suggested by the results of factor analysis, both of these composite variables were essentially unidimensional. Instructional practice-related activities in math lessons had a Cronbach's $\alpha$ of .55 and Content-related activities in math lessons had a Cronbach's $\alpha$ of .60. Details of item factor pattern coefficients can be found in Appendix D.

Five composite variables (i.e., ready to teach number, algebra, measurement, geometry, and data) were created from a series of 18 items related to teachers' reported level of readiness to teach each of the math topics. Specifically, Ready to teach number was measured by two items and each of the remaining composite variables by four items. Results of factor analyses suggested that these composite variables were essentially unidimesional. Factor pattern coefficients of all the items were greater than .30 and the Cronbach's $\alpha$ for each of the composite variables was relatively high, from .71 to .86. Specific details of item factor pattern coefficients can be found in Appendix D.

The next composite variable, professional development, was measured by five items related to various types of teacher training in the area of math instruction and assessment. Based on an examination of the scree plot, eigenvalues, factor pattern coefficients greater than .30 and interpretability of items, a single factor was retained. All five items were included in the calculation of the composite variable, professional development, which had a Cronbach's $\alpha$ of .78. A summary of item factor pattern coefficients for this composite variable can be found in Appendix D.

The composite variable, school resources for math instruction, was constructed from10 items related to availability of various school resources (e.g., textbook, supplies, computer, reference materials, and physical spaces and conditions) for math instruction. Through an examination of the scree plot, eigenvalues, and interpretability of variables, a single factor was retained. All the items were included in the computation of the composite variable, which had a Cronbach's $\alpha$ of .92. Details of item factor pattern coefficients can be found in Appendix D.

Finally, a factor analysis was conducted for six items related to the extent to which student factors (e.g., differences in academic abilities, background, and special needs) could affect math instruction. Results of the analysis suggested a single factor to be retained. All the items had factor pattern coefficients larger than .30 and thus were included in a composite variable labeled teacher's perception of math instructional limitations due to student factors. Cronbach's $\alpha$ for this composite variable was .81. A summary of item factor pattern coefficients for this composite variable can be found in Appendix D.

### *Content Experts' Validation of the Selected Variables*

In order to ensure that the selected variables for this study were important variables and that the way these variables were defined was in line with the view of teachers, educators, and content experts in the field of math education, the researcher set up two personal interviews with two professors of mathematics education at the College of Education, University of South Florida. A brief summary of these interviews are detailed below.

#### *Interview with Content Expert One*

The interview with Content Expert One who was an associate professor in math education at the College of Education, University of South Florida revolved around the topic of how students' math achievement was defined in the United States in particular, and whether such a definition was universally accepted. Content Expert One pointed out that math achievement could be represented by both overall math and separate content areas of math. Universally, however, math achievement had been frequently referred to as overall math. This was largely because in most research studies related to student math achievement, students were often asked about their confidence in learning overall math, and not in separate content areas of math. In terms of variables that were perceived to be positively related to math achievement, Content Expert One commented that she did not know of any variables that had consistent relationships with math achievement. However, recently, educators and educational policy makers called for more research concerning the relationships of opportunity to learn, students' self-confidence, time on math homework, class size, instructional time, activities in math lessons, teacher majoring in math, and teachers with math pedagogical skills. Content Expert One also added that there were several variables that showed little relationships with math achievement. These included: having a desk or a dictionary at home, valuing math, and tutoring (Content Expert One, personal interview, May 2, 2007).

*Interview with Content Expert Two*

The same questions that were asked in the interview with Content Expert One were used again in the interview with Content Expert Two who was another associate professor in Math Education at the College of Education, University of South Florida. Content Expert Two stated that, generally, math educators across countries agreed that there were five important content areas of math: number, measurement, algebra, geometry, and data. Although globally math achievement was often referred to as overall math, students did perform differently across content domains. Thus, math achievement should be reported by both overall math and by separate content areas. In response to the question regarding relationships of contextual variables and math achievement, Content Expert Two asserted that she had read a lot of research studies about math achievement but could not recall seeing any good predictor of math achievement. She could not recall any relationship between math scores and a variable that showed a consistent pattern across time or contexts. Because it was impossible to tell exactly or even the range of these relationships there was no clear guidance in terms of setting up hypotheses. Currently, math educators focus on investigating the impact of instructional practices (or opportunity to learn) and how well math lessons are delivered in terms of content knowledge, math teaching pedagogy, understanding of students' needs, teaching preparation in terms of number of math courses taken in undergraduate, and time on professional development on student math achievement (Content Expert Two, personal interview, May 4, 2007).

*Follow-up Interviews with Content Experts*

Prior to analyzing data for this dissertation, the researcher followed up with both Content Experts One and Two to obtain their opinions regarding the final list of variables for this study. Before the interviews, both content experts were provided with the list of variables and two questions: (1) Are these variables appropriate to include in the context of this study? and (2) Are there any variables not in the list but should be included? If yes, what are they and why?

The interviews took place in summer 2008. Consistently, both content experts indicated that the study had included important variables that were required to address the purpose of the study. However, several changes were suggested in order to improve the study from the mathematics content perspective.

1) Activities in math lessons would be better measured by two composite variables: Content-related activities in math lessons (four items) and Instructional practice-related activities in math lessons (five items).
2) Ready to teach overall math should be re-created to make five composite variables that reflect five content domains. Specifically, Ready to teach number measured by two items, Ready to teach algebra measured by four items, Ready to teach measurement measured by four items, Ready to teach geometry measured by four items, and Ready to teach data measured by four items.

3) Change the name of the variable, Professional development to Math-related professional development.
4) Change the name of the variable, Preparation to teach to Preparation to teach math content.
5) Opportunity to learn overall math should be re-created to make five composite variables that reflect five math content domains. Specifically, Opportunity to learn number measured by 10 items, opportunity to learn algebra measured by six items, opportunity to learn measurement measured by eight items, opportunity to learn geometry measured by 13 items, and opportunity to learn data measured by eight items.

As a result of these follow-up interviews, changes to the final list of variables were made as suggested by the content experts. Details of the variables listed in Table 6 reflect these changes.

## Data Analysis

### *Secondary Data Analysis*

The use of secondary data for research has become more common among social and behavioral science researchers. However, the choice to use secondary data must be made with consideration for its advantages and disadvantages (Rosenberg et al., 2006).

#### *Advantages*

The primary advantage of using secondary data for research is the conservation of time and expense because researchers can eliminate several steps in the research process, such as development of the measurement instruments, obtaining a research sample, the collection of the data, and the preparation of data for analysis by statistical packages

(Rosenberg et al., 2006). This is especially true when the target population for the research is national or international and the research questions require large sample sizes in order to obtain the power needed to make generalizations. More importantly, many of these large-scale secondary databases have high quality and are publicly accessible because they are maintained by well-established governmental organizations such as the National Center for Education Statistics (NCES), the National Science Foundation (NSF), and the Association for Institutional Research (AIR).

Another reason for researchers to make use of the excellent sources of existing large-scale secondary databases relates to the timeliness and richness of information provided in these databases (Martin, 2005). Researchers can easily find an educational longitudinal database or a trend database that includes data on more than a thousand variables collected over years. Thus, using the same database, researchers can conduct different studies to answer different research questions (Kiecolt & Nathan, 1985). Also, through archived secondary data sources, researchers can conduct re-analysis studies of past secondary databases using more advanced statistical methods in order to validate research findings produced by previous studies (Baker, 1992).

In terms of implications, from a theoretical perspective, the great level of variance in large-scale databases makes it possible for the testing of competing theoretical frameworks and revision of hypotheses (Kiecolt & Nathan, 1985). From a measurement perspective, secondary data analysis can support the refinement and improvement of the instruments through reliability analysis or confirmatory factor analysis (Hilton, 1992). Finally, from a methodological perspective, large-scale secondary data can be useful for delineating and developing new and improved statistical methods to solve pending research areas of concern (Kiecolt & Nathan, 1985).

*Disadvantages*

Despite the many advantages of using secondary data in research, there are a number of drawbacks associated with secondary data analysis that are important for researchers to consider. One of the major limitations of secondary data analysis is that the data are often collected for purposes other than the purpose of the secondary analysis (Gonzales, 2001). Additionally, although many existing secondary databases are publicly available for researchers to use, the time and costs associated with learning about the secondary data source including survey items, data structure, and technical documentation can be excessive (Hofferth, 2005). As pointed out by Gonzales (2001), Hahs-Vaughn (2005), and Moriarty et al. (1999), the two reasons that tend to keep researchers from using secondary data include the complex design of large-scale secondary data (e.g., multiple stage, clustering, and unequal sampling probabilities with non-uniform sampling weights) and the lack of sufficient skills that are required to effectively manage, manipulate and analyze large-scale data. Another limitation relates to data quality, including missing data, how latent constructs are defined, and the quality of supporting documentation for the data source (Rosenberg, Greenfield, & Dimick, 2006). Last but not least is the problem pertaining to the availability and accessibility of advanced statistical methodologies and specialized statistical software that must be used for analysis of complex, large-scale secondary data (Gonzales, 2001). As Gonzales (2001) put it:

> The application of sampling weights in simple statistical analyses is well understood, and methods for correctly estimating standard errors from clustered samples are increasingly available and acknowledged. In the case of more complex or more novel analytical methods, however, such as hierarchical modeling, the use of weights is difficult and not well understood, and software is not well developed in this respect. (Gonzales, 2001, p. 93)

### *Hierarchical Linear Modeling*

Hierarchical data structures are very common in the social and behavior sciences. A hierarchy consists of lower-level observations nested within higher-level(s) (Kreft & Leeuw, 2004). Examples include students nested within classes and classes nested within schools, or residents nested within neighborhoods, or repeated measurements nested within persons. Because of such naturally occurring clusters, researchers often collect data on variables at both the lower-level and the higher-level(s) of the hierarchy. For instance, in the TIMSS 2003 data, there are variables describing students (e.g., gender, self-confidence, attitudes towards learning, etc.), as well as variables describing schools (e.g., type of school, school size, school resources, etc.). Multilevel models are developed for analyzing hierarchically structured data (Kreft & Leeuw, 2004). The primary purpose of multilevel models is to capture the specific relationship between the lower-level and the higher-level(s) variables and the outcome variable. In the following sections, there is a discussion of the major advantages of using multilevel models to analyze hierarchically structured data over the traditional statistical regression models as well as theoretical and statistical assumptions of multilevel models.

*Advantages of Multilevel Models*

One problem associated with traditional statistical approaches is related to analysis of data at the aggregate level. Some researchers tend to collect data from individuals and then aggregate the data to gain insights into the groups to which those individuals belong. This approach is technically flawed because inferences about groups are incorrectly drawn from individual-level information (Luke, 2004). In other cases, data collected at the group level are disaggregated to the individual level. This approach is problematic because by ignoring group information, the model violates the independence of observations assumption leading to misestimated standard errors (standard errors are smaller than they should be).

In multilevel models, predictor variables are conceptually defined at different levels and the hypothesized relations between these predictor variables operate across different levels (Luke, 2004).Thus, unlike conventional regression approaches, the data in multilevel models can be analyzed in the context of the level and in relation to the other levels (i.e., within and between groups). It is essential to realize, however, that in multilevel models, characteristics or processes occurring at a higher level of analysis tend to influence characteristics or processes at a lower level (Luke, 2004).

Another advantage of using multilevel models over other traditional approaches addresses the issue of statistical or structural properties of the data. A major assumption of single-level, ordinary least squares (OLS) models is that the observations (and hence the error terms) are independent from one another. For hierarchically structured data, however, such an assumption is not valid because individuals who belong to the same group or context tend to have similar characteristics and thus, error terms tend to be correlated. If OLS is used for clustered data with correlated errors, the resulting standard errors are smaller than they should be, resulting in a greater chance of committing Type I errors. In contrast, by accounting for both within and between group variability at two or more levels simultaneously, multilevel models can estimate the appropriate, unbiased, errors (Luke, 2004). In addition, multilevel models allow for estimating cross-level effects that cannot be conceptually defined in conventional single-level regression models

Finally, unlike traditional statistical approaches where sample size must meet specific criteria, multilevel models are powerful in that they can handle relatively small sample size. Although the larger sample size will likely increase power of the study, multilevel models will be robust if the higher level sample size is at least 20 (Hox, 1995). According to simulation studies of Kreft (1996), there is adequate statistical power with 30 groups of 30 observations each; 60 groups with 25 observations each; and 150 groups with 5 observations each. These results indicate that the number of groups has more effect on statistical power in multilevel models than the number of observations, although both are important.

*Assumptions of Multilevel Models*

In order to ensure the validity of inferences based on the results of hierarchical linear models, the following assumptions pertaining to the adequacy of model specification and the consistency of parameter estimates in hierarchical linear models must be carefully tested (Raudenbush & Bryk, 2002):

1) Conditional on the level-1 variables, the within group errors ( $r_{ij}$ ) are normally distributed and independent with a mean of 0 in each group and equal variance across groups.

2) Any level-1 predictors of the outcome variable that are excluded from the model and thereby relegated to the error term ( $r_{ij}$ ) are independent of the level-1 variables that are included in the model (covariance equal 0).
3) In the random intercept only model, each group has a residual effect, $u_{0j}$. The distribution of these level-2 residual effects is normally distributed with mean 0 and variance $\tau_{00}$.
4) The effects of any excluded level-2 predictors from the model for the intercept are independent of other level-2 variables (covariance equal 0).
5) The level-1 error, $r_{ij}$, is independent of the level-2 residual effects, $u_{0j}$ (covariance equal 0).
6) Any level-1 predictors that are excluded from the level-1 model and as a result relegated to the error term, $r_{ij}$, are independent of the level-2 predictors in the model (covariance equal 0). In addition, any level-2 predictors that are excluded from the model and as a result relegated to the level-2 random effects, $u_{0j}$, are uncorrelated with the level-1 predictors (covariance equal 0).

Of these assumptions, assumptions 2, 4, and 6 focus on the relationship among the variables included in the structural part of the model (the level-1 and level-2 predictor variables) and those factors relegated to the error terms, $r_{ij}$ and $u_{0j}$. Misspecification of the model can cause bias in estimating level-1 and level-2 fixed effects. Assumptions 1, 3, and 5 are related to the random part of the model (i.e., $r_{ij}$ and $u_{0j}$). Their tenability affects the consistency of the estimates of the standard errors of level-2 fixed effects, the accuracy of the level-1 random effects, the variances for level-1 and level-2, and the accuracy of hypothesis tests and confidence intervals (Raudenbush & Bryk, 2002).

### *Analyses of TIMSS 2003 Database*

#### *Sampling Weights*

Because TIMSS 2003 utilizes a complex sampling design (see earlier section on sampling procedures), it is necessary to apply sampling weights when conducting analyses of the data in order to obtain unbiased population estimates (Martin, 2005). The sampling weights reflect the probability of selection of each school and student, taking into account any stratification or disproportional sampling of subgroups, and include adjustments for non-response (Martin, 2005).

Because the students within each country were selected using a probability sampling procedure, the probability of each student being selected as part of the sample is known. The sampling weight is the inverse of this selection probability. In a properly selected and weighted sample, the sum of the weights for the sample approximates the size of the population. In TIMSS 2003 study, each student's sampling weight, TOTWGT, is a composite of six factors: three weighting factors corresponding to the stages of the sampling design (i.e., school, class, and student), and three adjustment factors for non-participation at each of these stages. The use of TOTWGT ensures that the various subgroups that constitute the sample are properly and proportionally represented in the computation of population estimates (Martin, 2005).

When student and teacher data are to be analyzed together, the use of teacher weights are recommended. The math teacher weight, MATWGT, should be used to obtain estimates regarding students and their teachers. This weight is computed by dividing the sampling weight for the student by the number of math teachers that the student has (Martin, 2005). When conducting analyses at the school level, the use of the school weight is recommended. The school weight, SCHWGT, is the inverse of the probability of selection of the school, multiplied by its corresponding non-participation adjustment factor (Martin, 2005).

*Managing Multiple Databases*

Once four countries were selected, further managing and screening of the data were conducted in order to prepare the data for subsequent statistical analyses. Figure 1 displays a flow chart for this data management process. As indicated in Step 3 of the flowchart, the screening and sub-setting process of data management resulted in some reduction of sample sizes. For example, of the four countries included in this study, the United States is the only country where a second classroom was sampled in most schools. Therefore, in order to keep the data structures similar across countries, the researcher decided to keep only one classroom per school in this country. The classroom selected was the one with the most number of students. In addition, because there was a small number of schools in each country that had fewer than 10 students per school, a decision was made to eliminate these schools so that the final schools in the study had a minimum of 10 students.

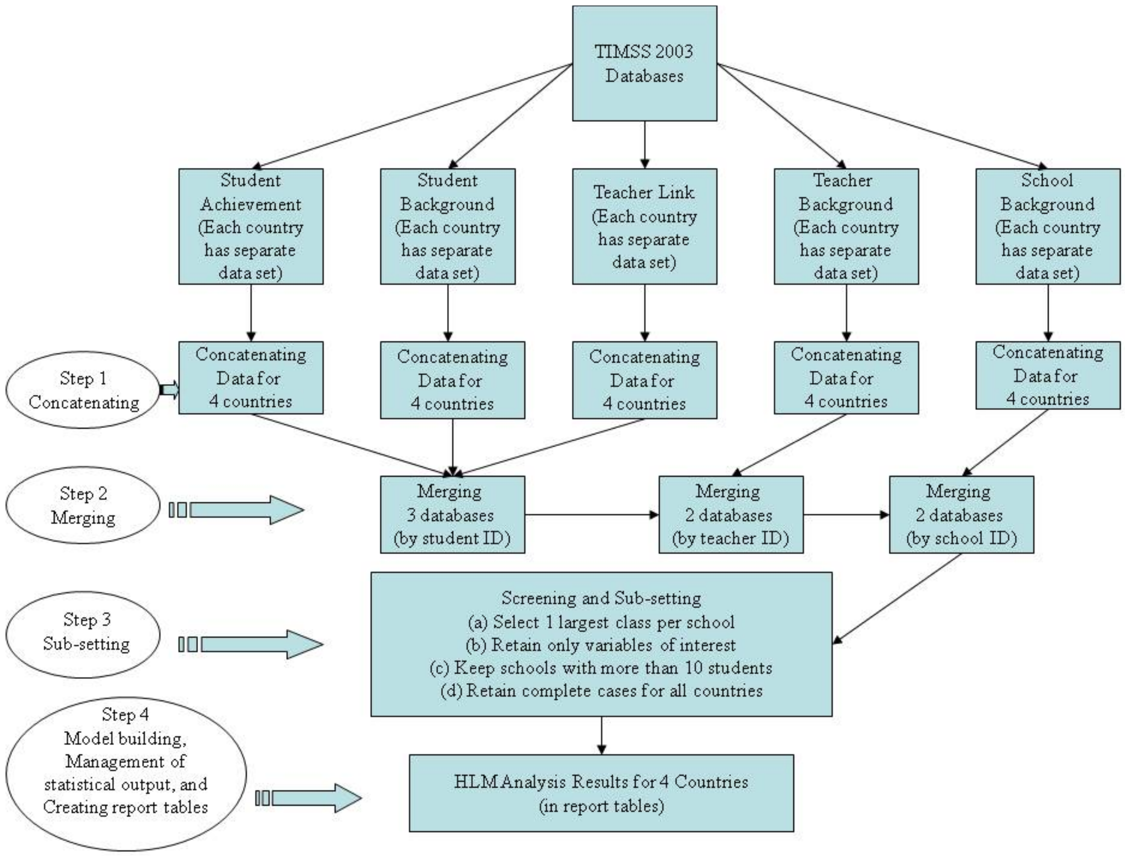



*Figure 1.* Flowchart for Managing Multiple Databases from TIMSS 2003

*Treatment of Missing Data*

Due to the complex and large-scale survey design of the TIMSS 2003, missing data were unavoidable and need to be addressed before any statistical analyses can be performed. In this study, an examination of missing data was conducted separately for each country at both the student level (level 1) and the classroom/school level (level 2). Next, listwise deletion as a missing data treatment method was employed to eliminate all the missing data at both level-1 and level-2. This step was conducted because in two-level HLM analyses, parameter estimates are computed based on complete cases.

Listwise deletion method was selected for this study because it is the simplest and most common method of handling missing data. Also, evidence from various research studies suggested that, in comparison with other methods of handling missing data such as pairwise deletion, listwise deletion tends to produce the least biased estimates (Allison, 2001; Roth, 1994). In a recent study of the impact of missing data in large-scale assessments, Phan and Kromrey (2006) found that statistical results produced from the listwise deletion method were comparable with those produced by the multiple imputations method, an increasingly promising method of missing data treatment (SAS, 2006; Mullis, 2001; Kromrey & Hines, 1994).

*Univariate Analysis*

Descriptive statistics such as weighted frequencies and weighted means were computed for the criterion and predictor variables by student level and school level by country using SAS v9.1.3 (SAS Institute Inc., 2005). In addition, figures and tables were used to display distributions of both criterion and predictor variables for each of the countries included in the study.

*Bivariate Analysis*

The bivariate relationships between level 1 predictor variables and level 2 predictor variables were also examined by individual country using SAS v9.1.3 (SAS Institute Inc., 2005).

*Hierarchical Linear Modeling Analysis*

Because the TIMSS 2003 data were reported by students nested in classes where there was one class sampled for each selected school, the analysis of the data was accomplished by the use of hierarchical linear modeling (HLM), a multi-level multiple regression technique useful in analyzing nested data (Raudenbush & Bryk, 2002). In order to proceed with HLM, the number of levels in the data needed to be specified and models needed to be constructed. The TIMSS 2003 data were best described in two levels: student level (level-1), and school and teacher level (level- 2). Level-1 was represented by student background and home resources variables which were unique across students. Level-2 was represented by instructional practices, teacher background and school background variables because each school had only one mathematics class sampled.

Although a third level might be incorporated in multi-level models, the analysis consisted of a series of two-level models for several reasons. First, the countries participated in the TIMSS 2003 voluntarily. No random selection procedure was applied at the country level. Thus, participating countries were not representative of countries in the world (TIMSS, 2003). Secondly, in TIMSS 2003, country background data were not collected (TIMSS, 2003). In addition, data on some variables of interest were not available for all of the countries included in this study because some countries opted not to administer certain survey items for country-specific reasons (TIMSS, 2003). Further, of the 50 participating countries, only four were selected to be included in this study. If a third level was to be incorporated, the small number of units of analysis at the third level (N = 4) would likely cause estimation problems (Raudenbush & Bryk, 2002). Finally, the chief purpose of the proposed study was to build models that would yield country-specific findings in terms of correlates of eighth-grade students' math achievement.

All subsequent HLM analyses were conducted using HLM 6, the specialized software developed for analysis of hierarchically structured data (Raudenbush, Bryk, Cheong, & Congdon, 2004). The chief reason that HLM 6 was selected for this study as opposed to other specialized software such as SAS Proc Mixed was that HLM 6 has the ability to incorporate appropriate complex design sampling weights at different levels of analysis (Raudenbush, Bryk, Cheong, & Congdon, 2004).

*Recoding Predictor Variables for HLM Analyses*

In order to improve interpretability of the results, both level-1 and level-2 predictors were recoded such that 0 represented the smallest category of a variable. For example, the predictor variable student self-confidence in learning math originally had three categories: 1= high, 2 = medium, and 3 = low. After recoding, the three categories of this variable were: 0 = low, 1 = medium, and 2 = high. There was one exception. The predictor, average number of math instructional hour per year was grand-mean centered.

*Models of the Study*

For each country, 23 models were constructed to represent level 1 and level 2 of the TIMSS 2003 data. The first model was the baseline or unconditional model which had no level 1 or level 2 variables. The regression equation is as follows.

$$Y_{ij} = \beta_{0j} + r_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$

In this model, $Y_{ij}$ is Mathematics score of student i in school j.

$\beta_{0j}$ is regression intercept of school j.

$\gamma_{00}$ is the overall average mathematics score for all schools.

$u_{0j}$ is the random effect of school j.

$r_{ij}$ is the random effect of student i in school j.

Each of the student background variables (i.e., gender, self confidence in learning math, student valuing of math, time on mathematics homework, and tutoring in math) then was entered separately in the unconditional model to make five level-1 models,

Models 2 to 6. Next, Model 7 was built to include all the student background variables. This model aimed to examine the extent to which student background variables were associated with math achievement. Similarly, Model 8 was constructed to examine the extent to which home resources for learning was related to student math performance. As an overall level-1 model, Model 9 was created to include all the student-related variables to the baseline model. The purpose of this model (Model 9) was to examine the relationship of each of the student-related variables in the presence of other variables and eighth-grade students' mathematics achievement. It is important to note that only those variables that were statistically significant in Model 9 were retained to include in level-2 models. The regression equations for nine level-1 models follow:

Model 2: $Y_{ij} = \beta_{0j} + \beta_{1j} Gender_{ij} + r_{ij}$

Model 3: $Y_{ij} = \beta_{0j} + \beta_{1j} Confidence_{ij} + r_{ij}$

Model 4: $Y_{ij} = \beta_{0j} + \beta_{1j} Valuing_{ij} + r_{ij}$

Model 5: $Y_{ij} = \beta_{0j} + \beta_{1j} TimeHomework_{ij} + r_{ij}$

Model 6: $Y_{ij} = \beta_{0j} + \beta_{1j} Tutoring_{ij} + r_{ij}$

Model 7:

$$Y_{ij} = \beta_{0j} + \beta_{1j} Gender_{ij} + \beta_{2j} Confidence_{ij} + \beta_{3j} Valuing_{ij} + \beta_{4j} TimeHomework_{ij}$$

$$+ \ \beta_{5j} Tutoring_{ij} + r_{ij}$$

$\beta_{pj} = \gamma_{p0} + u_{pj}$, where $p$ = 0, 1, 2, …5.

Model 8: $Y_{ij} = \beta_{0j} + \beta_{1j} HomeResources_{ij} + r_{ij}$

Model 9:

$$Y_{ij} = \beta_{0j} + \beta_{1j}Gender_{ij} + \beta_{2j}Confidence_{ij} + \beta_{3j}Valuing_{ij} + \beta_{4j}TimeHomework_{ij}$$

$$+ \beta_{5j}Tutoring_{ij} + \beta_{6j}HomeResources_{ij} + r_{ij}$$

$$\beta_{pj} = \gamma_{p0} + u_{pj}\text{, where } p = 0, 1, 2, \ldots 6.$$

In Model 2-9, $Y_{ij}$, $\beta_{0j}$, $\gamma_{00}$, $u_{0j}$, and $r_{ij}$ are as defined in the Baseline Model above.

$\beta_{1j}$ to $\beta_{6j}$ refer to regression slopes of school j

$\gamma_{p0}$ refer to the level 2 fixed effects

$u_{pj}$ refer to the level 2 random effects.

Similarly, at level-2, each of the instructional practices, teacher background and school background variables (i.e., opportunity to learn, homework assignment, activities in math lessons, instructional time, preparation to teach, ready to teach, professional development, class size, teacher perception and school resources) then were entered separately in Model 9. In addition, in order to estimate the amount of variance in math achievement that a set of variables (i.e., instructional practices, teacher background, school background, and all level-2 variables) account for, four combined models were constructed. Specifically, Models 10-14 represented instructional practices models, Models 15-18 represented teacher background models, Models 19-22 represented school background models, and finally, Model 23 represented the full model which included all level-2 variables and cross-level interaction terms that were statistically significant in earlier models. All level-2 models included random errors. The purpose of the level-2 models was to examine the relationship of instructional practices, teacher- and school-related factors as well as possible cross-level interactions of these variables and eighth-grade students' mathematics achievement. The regression equations for these models follow:

Model 10: $$\beta_{pj} = \gamma_{p0} + \gamma_{p1} OpportunityNu + \gamma_{p2} OpportunityAl + \gamma_{p3} OpportunityMe$$
$$+ \gamma_{p4} OpportunityGe + \gamma_{p5} OpportunityDa + u_{pj}$$

Model 11: $$\beta_{pj} = \gamma_{p0} + \gamma_{p1} HWAssignment + u_{pj}$$

Model 12: $$\beta_{pj} = \gamma_{p0} \gamma +_{p1} ActivitiesCon + \gamma_{p2} ActivitiesPra + u_{pj}$$

Model 13: $$\beta_{pj} = \gamma_{p0} + \gamma_{p1} InstructionalTime + u_{pj}$$

Model 14: $$\beta_{pj} = \gamma_{p0} + \gamma_{p1} OpportunityNu + \gamma_{p2} OpportunityAl + \gamma_{p3} OpportunityMe$$
$$+ \gamma_{p4} OpportunityGe + \gamma_{p5} OpportunityDa + \gamma_{p6} HWAssignment$$
$$+ \gamma_{p7} ActivitiesCon + \gamma_{p8} ActivitiesPra + \gamma_{p9} InstructionalTime + u_{pj}$$

Model 15: $$\beta_{pj} = \gamma_{p0} + \gamma_{p1} Preparation + u_{pj}$$

Model 16: $$\beta_{pj} = \gamma_{p0} + \gamma_{p1} ReadyNu + \gamma_{p2} ReadyAl + \gamma_{p3} ReadyMe$$
$$+ \gamma_{p4} ReadyGe + \gamma_{p5} ReadyDa + u_{pj}$$

Model 17: $$\beta_{pj} = \gamma_{p0} + \gamma_{p1} Development + u_{pj}$$

Model 18: $$\beta_{pj} = \gamma_{p0} + \gamma_{p1} Preparation + \gamma_{p2} ReadyNu + \gamma_{p3} ReadyAl + \gamma_{p4} ReadyMe$$
$$+ \gamma_{p5} ReadyGe + \gamma_{p6} ReadyDa + \gamma_{p7} Development + u_{pj}$$

Model 19: $$\beta_{pj} = \gamma_{p0} + \gamma_{p1} ClassSize + u_{pj}$$

Model 20: $$\beta_{pj} = \gamma_{p0} + \gamma_{p1} InstructionalLimit + u_{pj}$$

Model 21: $$\beta_{pj} = \gamma_{p0} + \gamma_{p1} SchoolResources + u_{pj}$$

Model 22:

$$\beta_{pj} = \gamma_{p0} + \gamma_{p1} ClassSize + \gamma_{p2} InstructionalLimit + \gamma_{p3} SchoolResources + u_{pj}$$

Model 23: $$\beta_{pj} = \gamma_{p0} + \gamma_{p1} OpportunityNu + \gamma_{p2} OpportunityAl + \gamma_{p3} OpportunityMe$$

$$+ \gamma_{p4} OpportunityGe + \gamma_{p5} OpportunityDa + \gamma_{p6} HWAssignment$$

$$+ \gamma_{p7} ActivitiesCon + \gamma_{p8} ActivitiesPra + \gamma_{p9} InstructionalTime$$

$$+ \gamma_{p10} Preparation + \gamma_{p11} ReadyNu + \gamma_{p12} ReadyAl + \gamma_{p13} ReadyMe$$

$$+ \gamma_{p14} ReadyGe + \gamma_{p15} ReadyDa + \gamma_{p16} Development$$

$$+ \gamma_{p17} ClassSize + \gamma_{p18} InstructionalLimit + \gamma_{p19} SchoolResources + u_{pj}$$

where $p$ = 0, 1, to 19.

In Model 10-23, $Y_{ij}$, $\beta_{0j}$, $\gamma_{00}$, $u_{0j}$, and $r_{ij}$ are as defined in the Baseline model above.

$\beta_{1j}$ to $\beta_{6j}$ and $u_{pj}$ are as defined in the Level 1 models.

$\gamma_{p0}$ to $\gamma_{p19}$ refer to the level 2 fixed effects.

*Power Analysis*

In HLM 2-level model analysis, the power of the study depends on the number of level-2 unit of analysis. Because the number of schools (level-2 units) in this study was relatively large, ranging from 52 to 271, there was enough power to detect the differences across schools (Kreft, 1996). Thus, no statistical power analysis was conducted.

*Summary*

In summary, Research Question 1 was addressed by using statistical results from Model 7 to make inferences about the extent to which student background variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, and tutoring in math) were associated with TIMSS 2003 eighth-grade math achievement in each country. As for Research Question 2, findings from Model 8 were used to make inferences about the extent to which home resources for learning variables (i.e., availability of calculator, computer, and desk for student use) were associated with TIMSS 2003 eighth-grade math achievement in each country.

In terms of Research Question 3, inferences about the extent to which instructional variables (i.e., opportunity to learn, activities in math lessons, amount of homework assignment, and instructional time) were associated with TIMSS 2003 eighth-grade math achievement in each country were made by using statistical results from Model 14. Similarly, Model 18 addressed Research Question 4 regarding the relationship between teacher-related variables (i.e., preparation to teach, ready to teach, and professional development) with TIMSS 2003 eighth-grade math achievement. Likewise, to address Research Question 5 concerning the association between school-related variables and TIMSS 2003 eighth-grade math achievement, statistical results from Model 22 were used. With statistical results obtained from the full model, Model 23, inferences were made about the extent to which all the statistically significant level-1 and level-2 variables as a set were related to TIMSS 2003 eighth-grade math achievement.

Finally, by visually and descriptively examining patterns of relationships between eighth-grade math achievement and contextual as well as background factors, this study identified important trends of relationships that tended to exist among developed and developing countries, as well as differences between these groups.

# CHAPTER FOUR

# RESULTS

## Results for the United States

### *Evaluation of Missing Data*

As a result of the listwise deletion method, the sample size for USA was reduced from 8808 students and 241 schools to 4,414 students and 153 schools. This means only 55.12% of the original sample had complete data on all variables of interest in this study. In order to evaluate the extent to which the data for USA were missing completely at random, the missingness on 19 level-2 variables were correlated. Results of this analysis suggested a non-randomness of missing data, with correlation coefficients ranging from .38 to .97 ($n$ = 153, $p$ <.001), indicating a modest to strong positive relationship among missingness indicators of the variables. In addition, when missingness was correlated with values of itself as well as values of other variables, only marginal correlations were observed ($r$ = -.20 to .19, $n$ = 153, $p$ = .005). In summary, the missing data mechanism for USA was not missing completely at random.

### *Univariate Analysis*

A descriptive examination of level-1 variables (i.e., overall math achievement, gender, self-confidence in learning math, valuing of math, time on math homework, extra math lessons, and home resources for learning math) was conducted using SAS 9.13. Of the complete sample of 4,414 eighth-grade students, 2,148 (48.66%) were female and 2,266 (51.34%) were male. On average, the weighted overall math achievement for USA students was 518.80 (*SD* = 72.43) with the lowest score of 274.96 and the highest score of 727.87 (see Table 7).

With regard to level-1 predictor variables, it appeared that, on average, eighth-grade students in the USA had most resources at home for learning math (*M* = 2.80, *SD* = .49), were above medium level of self-confidence in learning math (*M* = 1.34, *SD* = .78) and valuing of math (*M* = 1.52, *SD* = .64), spent little time on math homework (*M* = .74, *SD* = .57), and only had extra math lessons occasionally (*M* = .46, *SD* = .77) (see Table 7).

Table 7.

*Weighted Descriptive Statistics for Level-1 Variables for USA (N = 4,414)*

| Variable | *M* | *SD* | Min | Max |
|---|---|---|---|---|
| Overall math achievement | 518.80 | 72.43 | 274.96 | 727.87 |
| Self-confidence in learning math | 1.34 | 0.78 | 0 | 2 |
| Valuing of math | 1.52 | 0.64 | 0 | 2 |
| Time on math homework | 0.74 | 0.57 | 0 | 2 |
| Extra math lessons | 0.46 | 0.77 | 0 | 3 |
| Home resources for learning math | 2.80 | 0.49 | 0 | 3 |

Note: When weight was used to compute means in SAS, skewness and kurtosis were not produced

In terms of distributions of level-1 variables, the unweighted descriptive results from Table 8 suggested that all but two variables, extra math lessons and home resources for learning math, approximated normality, with skewness and kurtosis values within the range of -1.00 and 1.00.

Table 8.

*Unweighted Descriptive Statistics for Level-1 Variables for USA (N = 4,414)*

| Variable | *M* | *SD* | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| Overall math achievement | 516.95 | 73.12 | 274.96 | 727.87 | 0.03 | -0.31 |
| Self-confidence in learning math | 1.35 | 0.78 | 0 | 2 | -0.68 | -1.02 |
| Valuing of math | 1.52 | 0.64 | 0 | 2 | -0.98 | -0.13 |
| Time on math homework | 0.73 | 0.56 | 0 | 2 | 0.02 | -0.46 |
| Extra math lessons | 0.46 | 0.77 | 0 | 3 | 1.71 | 2.25 |
| Home resources for learning math | 2.78 | 0.5 | 0 | 3 | -2.44 | 6.24 |

Similarly, a descriptive analysis was conducted on 19 predictor variables at the school level. As can be seen from Table 8, on average, the percentage of students whose teachers reported their opportunity to learn math content domains (i.e., algebra, number, geometry, measurement, and data) were relatively high, ranging from 70.10 (*SD* = 26.89) for geometry to 99.65 (*SD* = 2.04) for number. Although not every math teacher was prepared to teach math content (*M* = .75, *SD* = .42), on average, they participated in various types of math-related professional development (*M* = 3.72, *SD* =1.62) and reported a high level of readiness to teach (*M* = 1.90, *SD* = .32 for measurement to *M* = 1.97, *SD* = .16 for number).

The data also suggested that in about half of the lessons, students were given activities related to math instructional practice (*M* = 1.76, *SD* = .25) and math content (*M* = 2.09, *SD* = .26). On average, a medium amount of homework was assigned to the students (*M* = 1.20, *SD* = .56). Finally, class size in USA schools tended to be small, less than 33 students (*M* = .57, *SD* = .67) and teachers' perception of instructional limitations due to student factors was low (*M*= .64, *SD* = .71). On average, the availability of school resources for math instruction was relatively high (*M* = 1.52, *SD* = .52). Noticeably, across 153 schools, the average math instructional hours per year varied greatly, ranging from 24.27 to 180, with a mean of 136.32 (*SD* = 28.52).

Of the 19 level-2 predictor variables, 10 approximated normal distributions with skewness and kurtosis values within the normality approximation range of -1.00 to 1.00. The nine variables that appeared to depart from normality included opportunity to learn number, measurement, and data; ready to teach number, algebra, measurement, geometry, and data; and average math instructional hours per year. specific skweness and kurtosis values for these variables can be found in Table 9.

Table 9.

*Unweighted Descriptive Statistics for Level-2 Variables for USA (N = 153)*

| Variable | *M* | *SD* | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| Opportunity to learn number | 99.65 | 2.04 | 80.00 | 100.00 | -7.44 | 62.55 |
| Opportunity to learn algebra | 78.86 | 23.20 | 16.67 | 100.00 | -0.67 | -0.78 |
| Opportunity to learn measurement | 83.71 | 19.89 | 0.00 | 100.00 | -1.69 | 3.58 |
| Opportunity to learn geometry | 70.10 | 26.89 | 0.00 | 100.00 | -0.71 | -0.30 |
| Opportunity to learn data | 83.96 | 21.65 | 0.00 | 100.00 | -1.70 | 3.05 |
| Amount of homework assignment | 1.20 | 0.56 | 0.00 | 2.00 | -0.17 | -0.16 |
| Instructional practice-related activities in math lessons | 1.76 | 0.25 | 1.04 | 2.33 | -0.07 | 0.08 |
| Content-related activities in math lessons | 2.09 | 0.26 | 1.33 | 2.74 | 0.04 | -0.02 |
| Preparation to teach | 0.75 | 0.42 | 0.00 | 1.00 | -1.17 | -0.59 |
| Ready to teach number | 1.97 | 0.16 | 1.00 | 2.00 | -6.00 | 34.43 |
| Ready to teach algebra | 1.91 | 0.28 | 1.00 | 2.00 | -2.94 | 6.87 |
| Ready to teach measurement | 1.90 | 0.32 | 0.00 | 2.00 | -3.21 | 10.67 |
| Ready to teach geometry | 1.91 | 0.32 | 0.00 | 2.00 | -4.16 | 18.82 |
| Ready to teach data | 1.92 | 0.27 | 1.00 | 2.00 | -3.07 | 7.61 |
| Math-related professional development | 3.72 | 1.62 | 0.00 | 5.00 | -1.08 | -0.10 |
| Class size for math instruction | 0.57 | 0.67 | 0.00 | 3.00 | 1.16 | 1.25 |
| School resources for math instruction | 1.52 | 0.52 | 0.00 | 2.00 | -0.38 | -1.16 |
| Teacher perceptions of math instructional limitation due to student factors | 0.64 | 0.71 | 0.00 | 2.00 | 0.73 | -0.69 |
| Average math instructional hours per year | 136.32 | 28.52 | 24.27 | 180.00 | -1.88 | 5.39 |

### *Bivariate Analysis*

An examination of bivariate relationships between variables was performed at each level. The results of weighted correlations among six level-1 variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, extra math lessons, and home resources for learning math) are presented in Appendix E. It appeared from these results that level-1 predictor variables were uncorrelated from each other. The magnitudes of the correlation coefficients ranged from .01 between valuing of math and time on math homework to .39 between self-confidence in learning math and valuing of math. It was interesting to note that gender tended to have a negative albeit weak relationship with all level-1 variables except for home resources for learning math ($r$ = .05).

At level-2, unweighted bivariate relationships were estimated for 19 predictor variables. The correlation matrix for these variables can be found in Appendix F. Unlike level-1, correlation coefficients of level-2 variables had a wider range, from -.31 between amount of homework assignment and teachers' perception of math instructional limitation due to student factors to .70 between ready to teach algebra and ready to teach measurement. As expected, correlation coefficients among the variables measuring the same construct tended to be stronger than those measuring different construct. For example, the correlations ranged from .19 to .45 for opportunity to learn variables, and .47 to .70 for ready to teach variables. Interestingly, it was observed that the number of math instructional hours per year had almost non-existent to very weak relationships with all of the level-2 variables ($r_s$ ranged from -.09 for ready to teach data to .15 for opportunity to learn measurement). Another interesting observation was that of 19 variables, 11 were found to have negative relationships ($r_s$ ranged from -02 to -.31) with teachers' perception of math instructional limitations due to student factors. These variables include all opportunity to learn variables, amount of homework assignment, activities in math lessons, and ready to teach geometry and school resources.

*Evaluation of HLM Assumptions*

In order to ensure tenability of results produced by multilevel models in this study, an evaluation of HLM assumptions through visual analysis of both level-1 and level-2 random effects of Model 14 was performed. Model 14 was selected because the results of HLM analysis suggested that it was the most efficient model to predict math achievement in the USA (see HLM Analysis for USA).

The data from Figure 2 suggested that level-1 residuals approximated a normal distribution. In terms of variance, the scatter plot between level-1 residuals and predicted math achievement, as illustrated in Figure 3, suggested that there was evidence of homogeneity of level-1 variance.

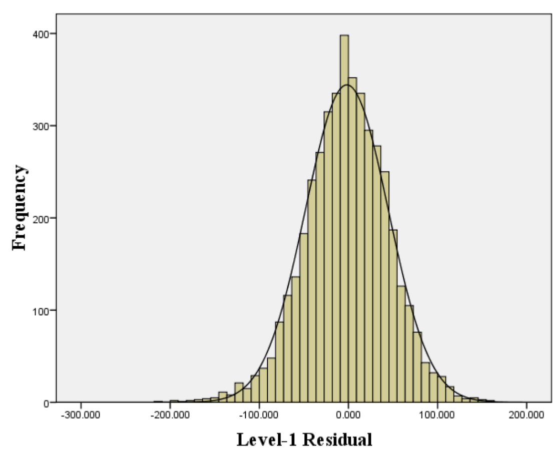



*Figure 2.* Histogram for Level-1 Residuals for USA

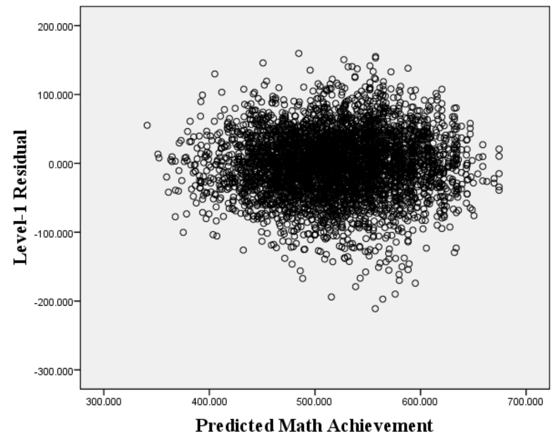



*Figure 3.* Level-1 Residuals by Predicted Math Achievement for USA

For level-2 random effects, the empirical Bayes residuals for the intercepts and slopes as well as empirical Bayes predicted math scores were used to construct the graphs in Figures 4-9. As can be seen from Figures 4-9, level-2 intercept residuals appeared to have a normal distribution and homogeneous variance.

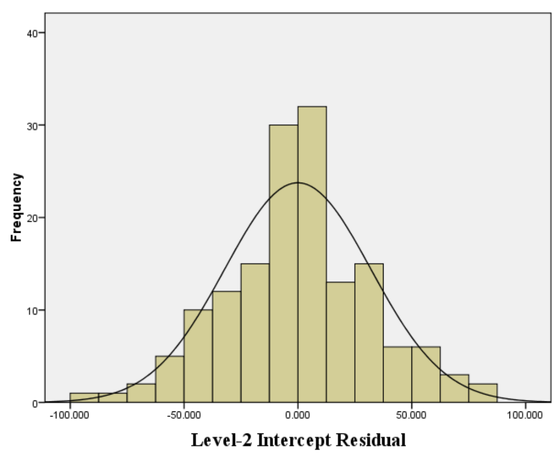



*Figure 4.* Histogram for Level-2 Intercept Residuals for USA

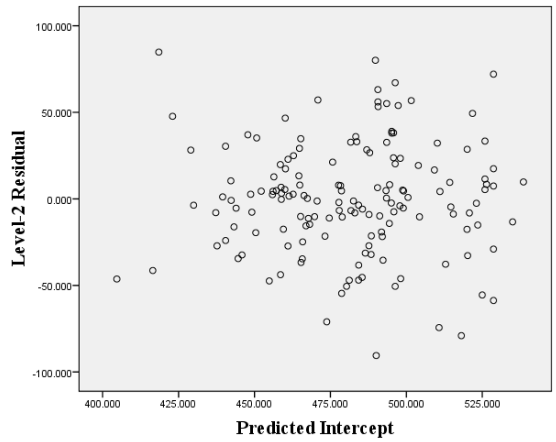



*Figure 5.* Level-2 Intercept Residuals by Predicted Intercept for USA

Similarly, Figure 6 suggests that level-2 residuals for the slope of Valuing of math approximated a normal distribution and Figure 7 provides evidence of homogeneity of variance.

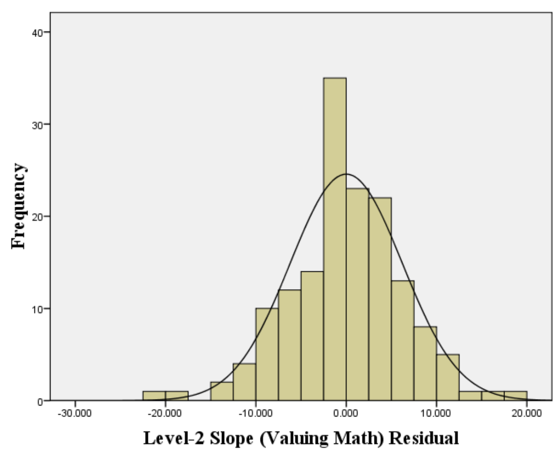



*Figure 6.* Histogram for Level-2 Slope (Valuing Math) Residuals for USA

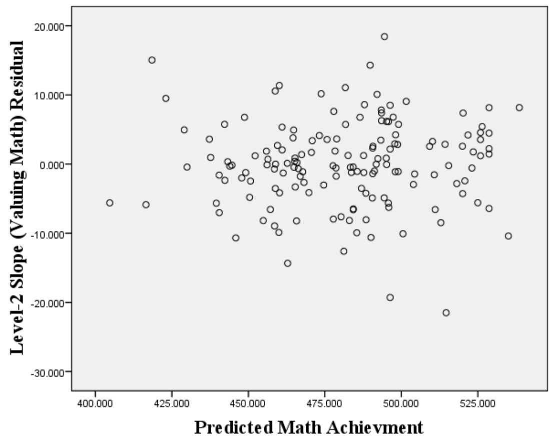



*Figure 7.* Level-2 Slope (Valuing Math) Residuals by Predicted Math Achievement for USA

Finally, although it appears from Figure 8 that level-2 residuals for the slope of Time on math homework had a slightly skewed distribution, their variances across school were relatively homogeneous (see Figure 9).

In summary, visual analyses of both level-1 and level-2 random effects suggested that the assumptions of normality and homogeneity of level-1 and level-2 random effects were satisfied.

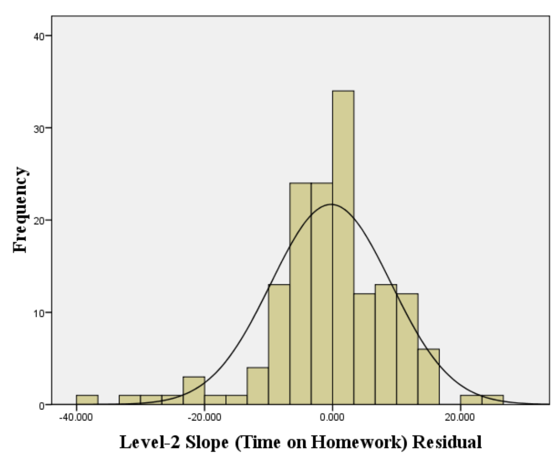



*Figure 8.* Histogram for Level-2 Slope (Time on Homework) Residuals for USA

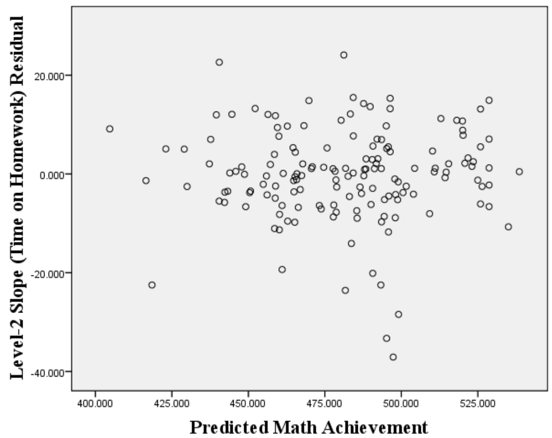



*Figure 9.* Level-2 Slope (Time on Homework) Residuals by Predicted Math Achievement for USA

*HLM Analysis*

*Unconditional model (Model 1)*

The HLM analysis started with the unconditional model where none of the level-1 or level-2 predictor was included in the model. The results of the unconditional model are presented in Table 10. For USA, the fixed effect for the intercept was 517.36 (*SE* = 4.78, *p* <.001). The average level of math achievement was significantly different across schools in the U.S ( $\tau_{00}$ = 2,703.82, *SE* = 52, *p* <.001). Within schools, the amount of unexplained variance was somewhat smaller than that between schools ( $\sigma^2$ = 2,591.61, *SE* = 50.91). The computed intra-class correlation (ICC) of .51 was substantial for the US, indicating a relatively strong level of natural clustering of students occurred between schools. In other words, approximately 51% of the total variance in math scores occurred between schools.

Table 10.
*Parameter Estimates for Unconditional Model for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 1 | Fixed | ICC | 0.51 | | |
| | | INT | 517.36 | 4.78 | <.001 |
| | Random | $\tau_{00}$ | 2703.82 | 52.00 | <.001 |
| | | $\sigma^2$ | 2591.61 | 50.91 | |

Note: ICC = intra-class correlation coefficient; INT = intercept

*Research Question 1*

To what extent are student background variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, and tutoring in math) associated with TIMSS 2003 eighth-grade math scores in each country?

In order to answer this research question, first, each of the student background variables was entered separately into Model 1 to predict math achievement. Then, as a group of variables, those that contributed significantly in Models 2-6 were included in Model 7 to predict math achievement. Finally, in order to evaluate model fit in terms of proportion of variance accounted for, a pseudo $R^2$ was computed for the current model against previously constructed models. Results of these models (Models 2-6) are presented in Table 11.

The data from Table 11 suggested that all of the fixed and random effects estimated by Models 2-6 were statistically significant, except for the fixed effect of time on homework in Model 5 ($\gamma$ = 2.38, *SE* = 2.45, *p* = .333). Interestingly, whereas self-confidence in learning math (Model 3) and valuing of math (Model 4) appeared to have positive relationships with math achievement ($\gamma$ = 28.24 and 11.66, *SE* = 1.29 and 1.74; *p* <.001 and .001, respectively), gender (Model 2) and extra math lessons (Model 6) appeared to have negative relationships with math achievement ( $\gamma$ = -6.47 and -19.36, *SE* = 2.04 and 1.22, *p* = .002 and .000, respectively).

An examination of pseudo $R^2$ across the five models (Models 2-6) suggested that the addition of individual predictors separately to the unconditional model (Model 1) to predict math achievement resulted in a reduction between 1% (Model 2) to 18% (Model 3) for the within school variance. For the between school variance, however, the amount of reduction was smaller, up to 7% (Model 3). In fact, in some models, the amount of between school variance slightly increased (for example, 1% for Model 1 and 2% for Model 6).

Table 11.

*Parameter Estimates for Models 2-6 (Level-1 Student Background) for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|
| 2 | Fixed | INT | 520.78 | 5.17 | <.001 | | |
| | | Gender | -6.47 | 2.04 | .002 | | |
| | Random | $\tau_{00}$ | 2957.66 | 54.38 | <.001 | | |
| | | Gender | 64.11 | 8.01 | .015 | | |
| | | $\sigma^2$ | 2564.59 | 50.64 | | | |
| | Pseudo $R^2$ | | | | | -0.01 | 0.01 |
| 3 | Fixed | INT | 479.63 | 4.44 | <.001 | | |
| | | Self-confidence | 28.24 | 1.29 | <.001 | | |
| | Random | $\tau_{00}$ | 2093.56 | 45.76 | <.001 | | |
| | | Self-confidence | 37.12 | 6.09 | .023 | | |
| | | $\sigma^2$ | 2133.91 | 46.19 | | | |
| | Pseudo $R^2$ | | | | | 0.07 | 0.18 |
| 4 | Fixed | INT | 499.82 | 4.84 | <.001 | | |
| | | Valuing math | 11.66 | 1.74 | <.001 | | |
| | Random | $\tau_{00}$ | 2071.39 | 45.51 | <.001 | | |
| | | Valuing math | 71.04 | 8.43 | .028 | | |
| | | $\sigma^2$ | 2513.45 | 50.13 | | | |
| | Pseudo $R^2$ | | | | | 0.03 | 0.03 |

Table 11.
*Parameter Estimates for Models 2-6 (Level-1 Student Background) for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|
| 5 | Fixed | INT | 515.18 | 5.21 | <.001 | | |
| | | Homework time | 2.38 | 2.45 | .333 | | |
| | Random | $\tau_{00}$ | 3030.85 | 55.05 | <.001 | | |
| | | Homework time | 330.89 | 18.19 | <.001 | | |
| | | $\sigma^2$ | 2503.19 | 50.03 | | | |
| | Pseudo $R^2$ | | | | | 0.00 | 0.03 |
| 6 | Fixed | INT | 526.47 | 4.68 | <.001 | | |
| | | Extra lessons | -19.36 | 1.22 | <.001 | | |
| | Random | $\tau_{00}$ | 2520.21 | 50.20 | <.001 | | |
| | | Extra lessons | 14.48 | 3.80 | .035 | | |
| | | $\sigma^2$ | 2389.20 | 48.88 | | | |
| | Pseudo $R^2$ | | | | | -0.02 | 0.08 |

Note: Pseudo $R^2$ refers to the difference in proportion of variance accounted for between the current models (Models 2-6) and the unconditional model (Model 1).

As a next step of model building, all of the student background variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, and tutoring in math) were included in the combined model, Model 7, to predict math achievement. Interestingly, in the presence of other variables in the model, only two out of five predictors had statistically significant fixed effects. With fixed effect of -15.22 (*SE* = 1.50, *p* <.001) for extra math lesson, it could be inferred that for each unit increase in extra math lesson (i.e., from 0 for never to 3 for daily), the students were expected to reduce 15.22 points in their math scores while controlling for other predictors in the model. Similarly, with fixed effect of 26.39 (*SE* = 1.24, *p* <.001) for self-confidence in learning math, it could be interpreted that for each unit increase in level of self-confidence in learning math (i.e., from 0 for low to 2 for high), it was expected that the students would improve 26.39 points in their math scores while controlling for other predictors in the model.

In terms of random effects, all were found statistically significant, except for those of gender ($\tau = 24.64$, *SE* = 4.96, $p = .500$) and extra math lessons ($\tau = 13.43$, *SE* = 3.67, $p = .078$). With the variance for the intercept of 2,142.83 (*SE* = 46.29, $p < .001$), it could be inferred that statistically significant differences existed across the school means of math achievement after adjusting for the five student background variables in the model. Similarly, it could be inferred that schools varied significantly in the relationships between math achievement and student self-confidence in learning math ($\tau = 23.54$, *SE* = 4.85, $p = .021$), student valuing of math ($\tau = 95.32$, *SE* = 9.76, $p < .001$), and time student spent on home work ($\tau = 273.02$, *SE* = 16.52, $p < .001$).

Table 12.
*Parameter Estimates for Models 7 (Level-1 Student Background) for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 7 | Fixed | INT | 492.91 | 5.24 | <.001 |
| | | Gender | -1.48 | 1.50 | .328 |
| | | Extra lessons | -15.22 | 1.05 | <.001 |
| | | Self-confidence | 26.39 | 1.24 | <.001 |
| | | Valuing math | -0.16 | 1.76 | .927 |
| | | Homework time | -3.24 | 2.18 | .140 |
| | Random | $\tau_{00}$ | 2142.83 | 46.29 | <.001 |
| | | Gender | 24.64 | 4.96 | >.500 |
| | | Extra lessons | 13.43 | 3.67 | .078 |
| | | Self-confidence | 23.54 | 4.85 | .021 |
| | | Valuing math | 95.32 | 9.76 | <.001 |
| | | Homework time | 273.02 | 16.52 | <.001 |
| | | $\sigma^2$ | 1892.30 | 43.50 | |

An evaluation of model fit was also conducted between Model 7 and previously constructed models, Models 2-6. As expected, the inclusion of student background variables in Model 7 yielded a considerable reduction in amount of variance accounted for in math achievement within schools, from 11% to 26% (see Table 12). Between schools, the amount of variance reduction was also observed, from 1% to 6%, except for the comparison with Model 3 where the amount of between school variance appeared to increase by 3%. In sum, Model 7 was more efficient than earlier models in predicting math achievement in the U.S.

Table 13.
*Comparison of $R^2$ between Model 7 and Previously Constructed Models for USA*

| Previous Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 2 | 0.05 | 0.26 |
| 3 | -0.03 | 0.11 |
| 4 | 0.01 | 0.25 |
| 5 | 0.04 | 0.24 |
| 6 | 0.06 | 0.21 |

*Research Question 2*

To what extent are home resources variables (i.e., availability of calculator, computer, and desk for student use) associated with TIMSS 2003 eighth-grade math scores in each country?

When the level-1 predictor home resources was added to the unconditional model to predict math achievement, a reduction of 1% was observed in the within school variance and a reduction of 5% in the between school variance (see Table 14). In this model, home resources had a statistically significant relationship with math achievement ($\gamma$ = 9.74, *SE* = 1.86, $p$ <.001). This means that for every unit increase in home resources, math achievement was expected to increase by 9.74 points, while not controlling for other variables. In addition, with the random effect for home resources being statistically significant ($\tau$ = 35.59, *SE* = 5.97, $p$ = .015), it could be inferred that the relationship between home resources and math achievement differed significantly across schools in the U.S.

Table 14.

*Parameter Estimates for Level-1 Home Resources Model for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|
| 8 | Fixed | INT | 489.62 | 6.06 | <.001 | | |
| | | Home resources | 9.74 | 1.86 | <.001 | | |
| | Random | $\tau_{00}$ | 1356.76 | 36.83 | <.001 | | |
| | | Home resources | 35.59 | 5.97 | .015 | | |
| | | $\sigma^2$ | 2576.33 | 50.76 | | | |
| | Pseudo $R^2$ | | | | | 0.05 | 0.01 |

Note: Pseudo $R^2$ refers to the difference in the proportion variance between Model 8 and Model 1.

Given the findings obtained from Models 7 and 8, five out of six student-related variables were entered into the unconditional model to make Model 9. Gender was excluded from Model 9 because both of its fixed and random effects were not statistically significant in Model 7. Also, in Model 9, the slope variance of extra math lessons was set to 0 because it was not statistically significant in Model 7.

As can be seen from Table 15, with the presence of other predictors in Model 9, only extra math lessons, self-confidence in learning math, and home resources had statistically significant fixed effects on math achievement. Specifically, whereas self-confidence in learning math ($\gamma$ = 26.31, *SE* = 1.25, *p* <.001) and home resources ($\gamma$ = 3.99, *SE* = 1.66, *p* = .017) were positively related to math achievement; an inverse relationship was observed between math achievement and extra math lessons ($\gamma$ = -15.63, *SE* = 1.07, *p* <.001). This could be interpreted to mean that the more learning resources students had at home and the more self-confidence they expressed in learning math, the higher were their math scores. However, it appears that the more frequently students took extra math lessons, the poorer math scores they achieved.

In terms of random effects, all were found statistically significant, except for the slope variance of home resources ($\tau$ = 28.29, *SE* = 5.32, *p* = .022). This suggests that the relationships between level-1 predictors (i.e., extra math lessons, self-confidence, and home resources) and math achievement varied significantly across schools.

As compared to Model 7, Model 9 appeared more efficient in that it accounted for more variance between schools (2%), even though no improvement in the variance within schools was noted. Compared to Model 8, Model 9 accounted for a significantly higher amount (26%) of the variance within school and a modest amount (2%) of the variance between schools. As a result of these comparisons, Model 9 was selected as the foundational level-1 model for further examination of the relationships between level-2 predictors and math achievement.

Table 15.
*Parameter Estimates for Combined Level-1 Predictors Model for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* | Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|
| 9 | Fixed | INT | 481.69 | 6.72 | <.001 | | | |
| | | Extra lessons | -15.63 | 1.07 | <.001 | | | |
| | | Self-confidence | 26.31 | 1.25 | <.001 | | | |
| | | Valuing math | -0.58 | 1.72 | .736 | | | |
| | | Homework time | -3.17 | 2.14 | .142 | | | |
| | | Home resources | 3.99 | 1.66 | .017 | | | |
| | Random | $\tau_{00}$ | 1874.76 | 43.30 | <.001 | | | |
| | | Self-confidence | 29.61 | 5.44 | .045 | | | |
| | | Valuing math | 92.34 | 9.61 | .014 | | | |
| | | Homework time | 252.02 | 15.87 | <.001 | | | |
| | | Home resources | 28.29 | 5.32 | .223 | | | |
| | | $\sigma^2$ | 1898.74 | 43.57 | | | | |
| | Pseudo $R^2$ | | | | | 7 | 0.02 | 0.00 |
| | | | | | | 8 | 0.02 | 0.26 |

Note: Pseudo $R^2$ refers to the difference in the proportion of variance between Models 7 and 8 and Model 9.

*Research Question 3*

To what extent are instructional variables (i.e., opportunity to learn, activities in math lessons, amount of homework assignment, and instructional time) associated with TIMSS 2003 eighth-grade math scores in each country?

In addressing this research question, a similar strategy for model building used in Research Question 1 was applied. That is, each of the level-2 instructional practice variables was first added to the foundational level-1 model (Model 9) to make Models 10-13. Then, as a group, those variables with significant fixed effects in Models 10-13 were included in the combined instructional practice model, Model 14. It is important to note that in these models, the slope variance of home resources was set to 0 because it was not statistically significant in Model 9. Also, all possible cross level interactions between level-1 and level-2 predictors were allowed in these models. The results of Models 10-14 are presented in Tables 16-18.

As can be seen in Table 16, Model 10 with opportunity to learn math topics as level-2 predictors of math achievement yielded three statistically significant cross-level interactions: (1) opportunity to learn data by self-confidence in learning math ($\gamma$ = -.16, *SE* = .07, *p* = .018), (2) opportunity to learn geometry by time student spent on homework ($\gamma$ = -.22, *SE* = .08, *p* = .006), and (3) opportunity to learn measurement by time student spent on homework ($\gamma$ = .23, *SE* = .11, *p* = .042).

Table 16 also showed that when amount of homework assignment, content-related activities and instructional practice-related activities in math lessons, and average number of math instructional hours per year were added to Models 11-13, no statistically significant cross- level interaction effects were detected. Of the level-2 main effects, only homework assignment was found statistically significant in Model 11 ($\gamma$ = 37.04, *SE* = 7.41, *p* <.001). This means that with every unit increase in amount of homework assignment, students' math scores were expected to increase by 37.04 points after adjusting for level-1 variables but not for other level-2 variables in the model.

Table 16.

*Parameter Estimates for Level-2 Instructional Practices Models for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 10 | Fixed | INT | 390.22 | 143.87 | .008 |
| | | Opportunity_algebra | 0.34 | 0.19 | .076 |
| | | Opportunity_data | -0.33 | 0.26 | .219 |
| | | Opportunity_geometry | 0.19 | 0.20 | .350 |
| | | Opportunity_measurement | 0.24 | 0.24 | .322 |
| | | Opportunity_number | 0.60 | 1.46 | .678 |
| | | Extra lessons | -15.53 | 1.05 | <.001 |
| | | Self-confidence | 26.69 | 25.23 | .292 |
| | | Opportunity_algebra*Self-confidence | 0.11 | 0.06 | .061 |
| | | Opportunity_data*Self-confidence | -0.16 | 0.07 | .018 |
| | | Opportunity_geometry*Self-confidence | 0.04 | 0.07 | .529 |
| | | Opportunity_measurement*Self-confidence | -0.08 | 0.08 | .345 |
| | | Opportunity_number*Self-confidence | 0.07 | 0.28 | .794 |
| | | Valuing math | -14.18 | 36.67 | .699 |
| | | Opportunity_algebra*Valuing math | 0.00 | 0.09 | .976 |
| | | Opportunity_data*Valuing math | 0.05 | 0.10 | .573 |
| | | Opportunity_geometry*Valuing math | -0.06 | 0.06 | .336 |
| | | Opportunity_measurement*Valuing math | 0.15 | 0.09 | .071 |
| | | Opportunity_number*Valuing math | 0.01 | 0.38 | .986 |
| | | Homework time | 20.61 | 41.85 | .623 |
| | | Opportunity_algebra*Homework time | -0.04 | 0.09 | .684 |
| | | Opportunity_data*Homework time | 0.18 | 0.13 | .184 |
| | | Opportunity_geometry*Homework time | -0.22 | 0.08 | .006 |
| | | Opportunity_measurement*Homework time | 0.23 | 0.11 | .042 |
| | | Opportunity_number*Homework time | -0.40 | 0.45 | .383 |
| | | Home resources | 3.44 | 1.58 | .029 |
| | Random | $\tau_{00}$ | 1819.39 | 42.65 | <.001 |
| | | Self-confidence | 19.27 | 4.39 | .073 |
| | | Valuing math | 92.47 | 9.62 | <.001 |
| | | Homework time | 213.47 | 14.61 | <.001 |
| | | $\sigma^2$ | 1909.71 | 43.70 | |

Table 16.

*Parameter Estimates for Level-2 Instructional Practices Models for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 11 | Fixed | INT | 437.57 | 9.61 | <.001 |
| | | Homework assignment | 37.04 | 7.41 | <.001 |
| | | Extra lessons | -15.45 | 1.06 | <.001 |
| | | Self-confidence | 28.47 | 2.65 | <.001 |
| | | Homework assignment*Self-confidence | -1.77 | 2.33 | .448 |
| | | Valuing math | 1.48 | 3.28 | .652 |
| | | Homework assignment*Valuing math | -1.65 | 2.69 | .540 |
| | | Homework time | 0.42 | 5.31 | .938 |
| | | Homework assignment*Homework time | -2.75 | 4.08 | .501 |
| | | Home resources | 3.68 | 1.57 | .019 |
| | Random | $\tau_{00}$ | 1503.34 | 38.77 | <.001 |
| | | Self-confidence | 29.56 | 5.44 | .042 |
| | | Valuing math | 97.11 | 9.85 | <.001 |
| | | Homework time | 259.64 | 16.11 | <.001 |
| | | $\sigma^2$ | 1904.70 | 43.64 | |
| 12 | Fixed | INT | 446.20 | 44.68 | <.001 |
| | | Content_activities | 22.31 | 27.04 | .411 |
| | | Instruction_activities | -5.57 | 23.02 | .809 |
| | | Extra lessons | -15.68 | 1.06 | <.001 |
| | | Self-confidence | 15.95 | 10.77 | .140 |
| | | Content_activities*Self-confidence | -1.57 | 5.11 | .759 |
| | | Instruction_activities*Self-confidence | 7.71 | 5.57 | .168 |
| | | Valuing math | 10.90 | 16.26 | .504 |
| | | Content_activities*Valuing math | 1.37 | 7.52 | .856 |
| | | Instruction_activities*Valuing math | -8.07 | 6.65 | .227 |
| | | Homework time | -1.41 | 17.31 | .936 |
| | | Content_activities*Homework time | -1.05 | 8.84 | .906 |
| | | Instruction_activities*Homework time | 0.28 | 8.62 | .975 |
| | | Home resources | 3.67 | 1.58 | .020 |
| | Random | $\tau_{00}$ | 1935.48 | 43.99 | <.001 |
| | | Self-confidence | 29.50 | 5.43 | .043 |
| | | Valuing math | 91.41 | 9.56 | <.001 |
| | | Homework time | 257.30 | 16.04 | <.001 |
| | | $\sigma^2$ | 1906.97 | 43.67 | |
| 13 | Fixed | INT | 483.21 | 6.62 | <.001 |
| | | Instructional hours | 0.00 | 0.15 | .995 |
| | | Extra lessons | -15.69 | 1.06 | <.001 |

Table 16.
*Parameter Estimates for Level-2 Instructional Practices Models for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Self-confidence | 26.20 | 1.27 | <.001 |
| | | Instructional hours*Self-confidence | -0.04 | 0.04 | .342 |
| | | Valuing math | 0.08 | 1.65 | .964 |
| | | Instructional hours*Valuing math | 0.10 | 0.06 | .068 |
| | | Homework time | -3.08 | 2.17 | .158 |
| | | Instructional hours*Homework time | 0.04 | 0.04 | .398 |
| | | Home resources | 3.52 | 1.59 | .027 |
| | Random | $\tau_{00}$ | 1971.50 | 44.40 | <.001 |
| | | Self-confidence | 29.07 | 5.39 | .040 |
| | | Valuing math | 85.87 | 9.27 | .001 |
| | | Homework time | 254.53 | 15.95 | <.001 |
| | | $\sigma^2$ | 1906.11 | 43.66 | |

In terms of model fit, in comparison with the foundational level-1 model (Model 9), Model 10 appeared to be the most efficient model because the amount of explained variance between schools in this model increased by 19% (see Table 17). As for the within school variance, no significant difference was observed between Models 10-13 and Model 9 (pseudo $R^2$ = 0 in Models 11-13 and -.01 in Model 10).

Table 17.
*Comparison of $R^2$ between Level-2 Instructional Practice Models and Foundational Level-1 Model for USA*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 10 vs. 9 | 0.19 | -0.01 |
| 11 vs. 9 | 0.09 | 0.00 |
| 12 vs. 9 | 0.07 | 0.00 |
| 13 vs. 9 | 0.00 | 0.00 |

Similar to Model 10, when using all the level-2 instructional practice variables to predict math achievement, Model 14 produced three statistically significant cross-level interaction effects (see Table 18). First, opportunity to learn data interacted with self-confidence in learning math ($\gamma$ = -.17, *SE* = .06, *p* = .012). Second, opportunity to learn measurement interacted with student valuing of math ($\gamma$ = .16, *SE* = .08, *p* = .048). And, third, opportunity to learn geometry interacted with time student spent on homework ($\gamma$ = -.23, *SE* = .08, *p* = .005). Interestingly, in this model, all the random effects were statistically significant, except for student self-confidence in learning math (τ = 20.62, *SE* = 4.54, *p* = .074). This suggested that, in the U.S., the positive relationship between student self-confidence in learning math and math achievement were similar across schools.

Table 18.

*Parameter Estimates for the Combined Level-2 Instructional Practices Model for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | INT | 363.49 | 146.56 | .015 |
| | | Homework assignment | 37.80 | 7.59 | <.001 |
| | | Opportunity_algebra | 0.37 | 0.19 | .051 |
| | | Opportunity_data | -0.45 | 0.22 | .048 |
| | | Opportunity_geometry | 0.31 | 0.17 | .065 |
| | | Opportunity_measurement | 0.14 | 0.20 | .469 |
| | | Opportunity_number | 0.41 | 1.46 | .782 |
| | | Content_activities | 5.62 | 22.52 | .803 |
| | | Instruction_activities | -2.32 | 17.86 | .897 |
| | | Instructional hours | -0.03 | 0.14 | .830 |
| | | Extra lessons | -15.51 | 1.05 | <.001 |
| | | Self-confidence | 13.41 | 23.58 | .570 |
| | | Homework assignment*Self-confidence | -0.60 | 2.31 | .796 |
| | | Opportunity_algebra*Self-confidence | 0.12 | 0.06 | .057 |
| | | Opportunity_data*Self-confidence | -0.17 | 0.06 | .012 |
| | | Opportunity_geometry*Self-confidence | 0.05 | 0.07 | .515 |
| 14 | Fixed | Opportunity_measurement*Self-confidence | -0.07 | 0.08 | .384 |
| | | Opportunity_number*Self-confidence | 0.11 | 0.26 | .676 |
| | | Content_activities | -2.18 | 5.23 | .677 |
| | | Instruction_activities | 8.12 | 5.25 | .124 |
| | | Instructional hours | -0.04 | 0.03 | .134 |
| | | Valuing math | 7.79 | 39.72 | .845 |
| | | Homework assignment*Valuing math | -2.65 | 2.83 | .350 |
| | | Opportunity_algebra*Valuing math | 0.01 | 0.09 | .955 |
| | | Opportunity_data*Valuing math | 0.08 | 0.10 | .401 |
| | | Opportunity_geometry*Valuing math | -0.07 | 0.06 | .253 |
| | | Opportunity_measurement*Valuing math | 0.16 | 0.08 | .048 |
| | | Opportunity_number*Valuing math | -0.02 | 0.40 | .958 |
| | | Content_activities*Valuing math | -0.13 | 8.21 | .987 |
| | | Instruction_activities*Valuing math | -9.84 | 6.35 | .123 |
| | | Instructional hours*Valuing math | 0.08 | 0.05 | .149 |
| | | Homework time | 33.79 | 43.98 | .444 |

Table 18.
*Parameter Estimates for the Combined Level-2 Instructional Practices Model for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Homework assignment*Homework time | -4.73 | 3.96 | .235 |
| | | Opportunity_algebra*Homework time | -0.03 | 0.10 | .763 |
| | | Opportunity_data*Homework time | 0.19 | 0.14 | .166 |
| | | Opportunity_geometry*Homework time | -0.23 | 0.08 | .005 |
| | | Opportunity_measurement*Homework time | 0.25 | 0.13 | .050 |
| | | Opportunity_number*Homework time | -0.36 | 0.47 | .447 |
| | | Content_activities*Homework time | -3.64 | 8.49 | .669 |
| | | Instruction_activities*Homework time | -3.14 | 8.20 | .702 |
| | | Instructional hours*Homework time | 0.01 | 0.04 | .779 |
| | | Home resources | 3.52 | 1.55 | .024 |
| | | $\tau_{00}$ | 1408.20 | 37.53 | <.001 |
| | | Self-confidence | 20.62 | 4.54 | .074 |
| | Random | Valuing math | 90.99 | 9.54 | <.001 |
| | | Homework time | 224.41 | 14.98 | <.001 |
| | | $\sigma^2$ | 1907.62 | 43.68 | |

As evident in Table 19, compared to previously constructed models (Models 9-13), the amount of explained variance between schools in Model 14 was more significant. For example, an increase of 25% in the between school variance was observed when using Model 14 instead of Models 9 or 13. At minimum, changing from Models 10 to Model 14 would result in 7% more of the variance between schools to be accounted for. However, for the variance within schools, no change was noted across these models. Thus, in consideration of the amount of the explained variance between schools, Model 14 surpassed previously constructed models in predicting math achievement.

Table 19.
*Comparison of $R^2$ between Model 14 and Previously Constructed Models 9-13 for USA*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 14 vs. 9 | 0.25 | 0.00 |
| 14 vs. 10 | 0.07 | 0.00 |
| 14 vs. 11 | 0.17 | 0.00 |
| 14 vs. 12 | 0.19 | 0.00 |
| 14 vs. 13 | 0.25 | 0.00 |

The modeled means of predicted math achievement for the significant interactions resulted from Model 14 are displayed in Figures 10-12. It is worth noting that because the illustrations focused on the nature of the cross-level interactions, the vertical axis on the interaction plots was not scaled to the actual values of the predicted math scores.

The data in Figure 10 suggested that the relationship between students' reported valuing of math and math achievement was different across levels of opportunity to learn measurement as a math topic. Specifically, when opportunity to learn measurement was low, there was little difference in math achievement among students with low, medium and high levels of valuing math. However, as opportunity to learn measurement increased, the size of the difference in math achievement among these students got larger. As expected, students with higher level of valuing of math tended to achieve higher math scores than those with medium and low level of valuing math.

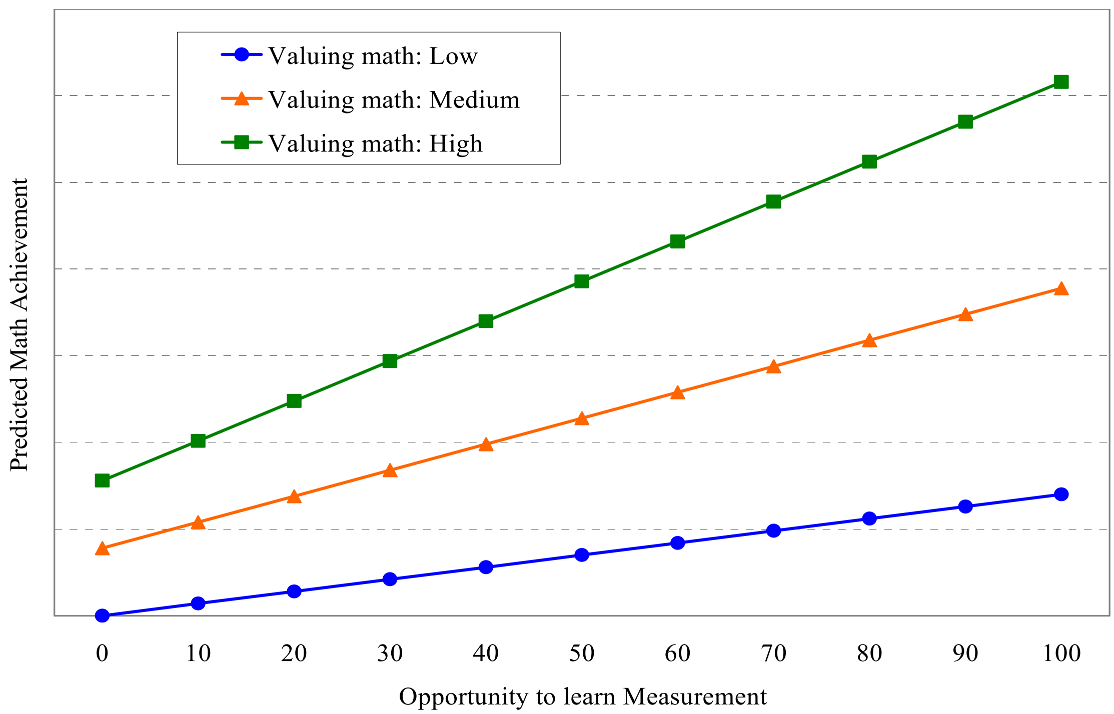



*Figure 10.* Interaction between Valuing of Math and Opportunity to Learn measurement for USA

The data in Figure 11 depict the interaction between time student spent on homework and opportunity to learn geometry. Interestingly, it was found that, for those students who reported spending a high amount of time on homework, their predicted math scores tended to be lower when they had more opportunity to learn geometry and higher when they had no or little opportunity to learn geometry. However, as students spent lesser amounts of time on homework, an inverse pattern of relationship between opportunity to learn geometry and math achievement was observed. That is, for these groups of students, no or little opportunity to learn geometry was associated with a lower math scores and high opportunity to learn geometry was associated with a higher math scores. It appears that, increased opportunity to learn geometry worked best for the group of students who reported spending a low amount of time on homework.

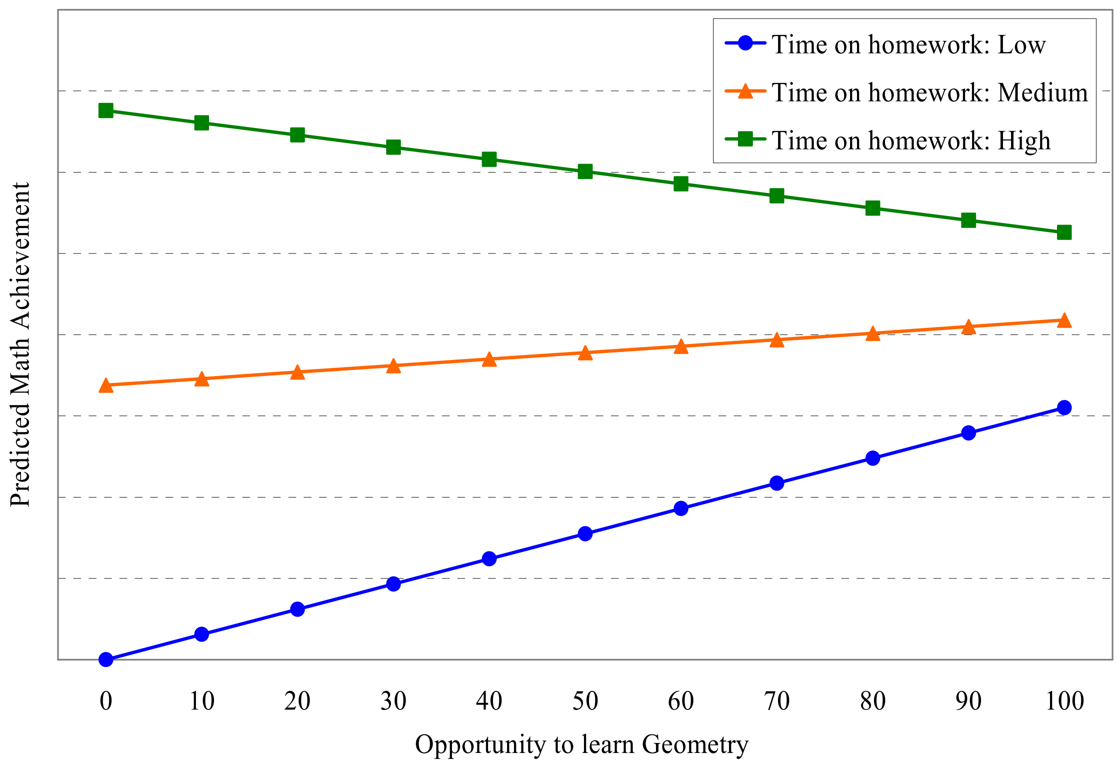



*Figure 11.* Interaction between Time Student Spent on Homework and Opportunity to Learn geometry for USA

The nature of the interaction between student self-confidence in learning math and opportunity to learn data is displayed in Figure 12. Surprisingly, it was noted that students tended to perform similarly low in math when there was a high opportunity to learn data. As opportunity to learn data decreased, students' math scores increased significantly, with students who reported having a high level of self-confidence in learning math tended to perform better than their peers who reported a lower level of self-confidence in learning math.

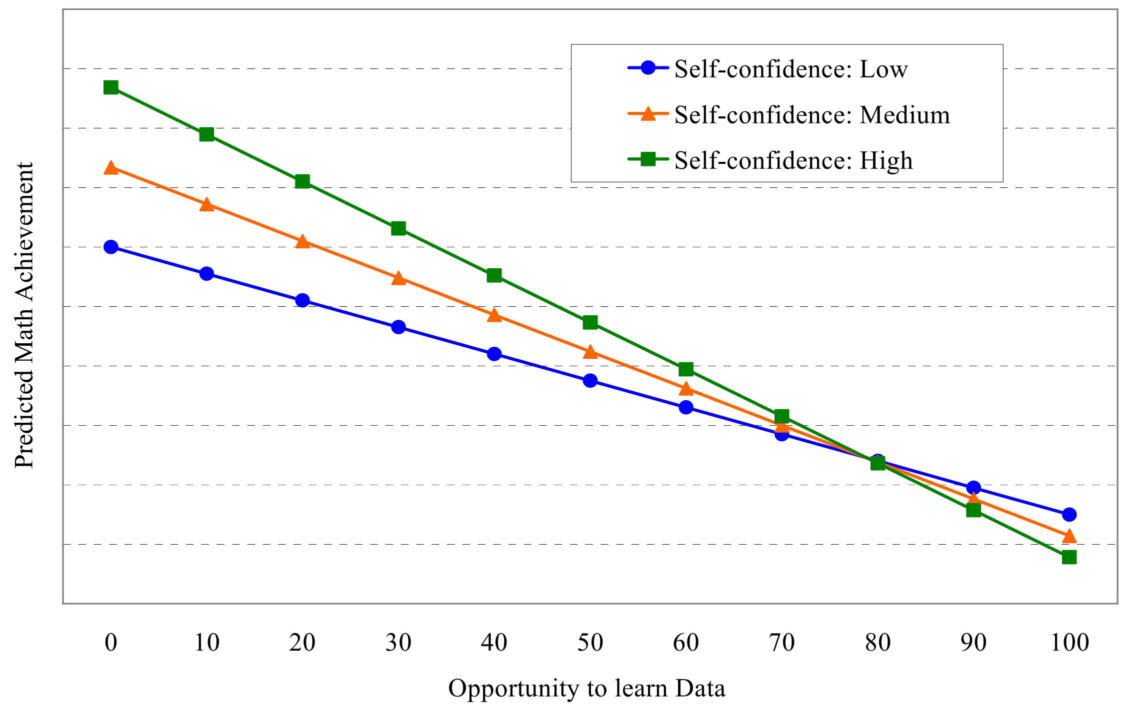



*Figure 12.* Interaction between Self-confidence in Learning Math and Opportunity to Learn data for USA

*Research Question 4*

To what extent are teacher-related variables (i.e., preparation to teach, ready to teach, and professional development) associated with TIMSS 2003 eighth-grade math scores in each country?

Similarly, incremental model building strategies were applied to examine the effects of teacher-related variables (i.e., preparation to teach, ready to teach, and professional development) on math achievement. Results of these models (Models 15-18) are presented in Table 20-23.

Interestingly, as shown in Table 20, preparation to teach (Model 15), ready to teach math topics (Model 16), and professional development (Model 17) as single level-2 predictors in the model did not appear to have statistically significant relationships with math achievement. In Model 16, however, there were three statistically significant cross-level interaction effects: (1) ready to teach algebra by student self-confidence in learning math ($\gamma$ = 8.06, *SE* = 3.60, *p* = .027); (2) ready to teach number by time student spent on homework ($\gamma$ = -.33.81, *SE* = 14.47, *p* = .021), and (3) ready to teach data and time student spent on homework ($\gamma$ = 15.46, *SE* = 6.61, *p* = .021).

In these models, all of the random effects were statistically significant, suggesting that a significant amount of variance in math achievement remained unexplained both within and between schools.

Table 20.
*Parameter Estimates for Teacher Background Models for USA*

| Model | Effect | Parameters | Estimates | *SE* | *P* |
|---|---|---|---|---|---|
| 15 | Fixed | INT | 484.02 | 11.48 | <.001 |
| | | Preparation | -1.81 | 11.74 | .878 |
| | | Extra lessons | -15.59 | 1.05 | <.001 |
| | | Self-confidence | 29.27 | 2.07 | <.001 |
| | | Preparation*Self-confidence | -4.08 | 2.63 | .123 |
| | | Valuing math | -4.77 | 4.12 | .250 |
| | | Preparation*Valuing math | 6.02 | 4.59 | .192 |
| | | Homework time | -6.60 | 3.59 | .068 |
| | | Preparation*Homework time | 4.71 | 4.44 | .290 |
| | | Home resources | 3.58 | 1.57 | .023 |
| | Random | $\tau_{00}$ | 1957.94 | 44.25 | <.001 |
| | | Self-confidence | 29.00 | 5.39 | .046 |
| | | Valuing math | 85.47 | 9.24 | .001 |
| | | Homework time | 248.97 | 15.78 | <.001 |
| | | $\sigma^2$ | 1906.63 | 43.66 | |
| 16 | Fixed | INT | 462.94 | 29.00 | <.001 |
| | | Ready_number | -32.06 | 34.82 | .359 |

Table 20.
*Parameter Estimates for Teacher Background Models for USA*

| Model | Effect | Parameters | Estimates | *SE* | *P* |
|---|---|---|---|---|---|
| | | Ready_algebra | -2.01 | 17.41 | .909 |
| | | Ready_measurement | 1.03 | 17.30 | .953 |
| | | Ready_geometry | 30.78 | 24.48 | .211 |
| | | Ready_data | 13.59 | 18.64 | .467 |
| | | Extra lessons | -15.64 | 1.05 | <.001 |
| | | Self-confidence | 41.48 | 11.43 | .001 |
| | | Ready_number*Self-confidence | -10.94 | 7.08 | .124 |
| | | Ready_algebra*Self-confidence | 8.06 | 3.60 | .027 |
| | | Ready_measurement*Self-confidence | -2.05 | 6.36 | .748 |
| | | Ready_geometry*Self-confidence | 1.62 | 2.70 | .549 |
| | | Ready_data*Self-confidence | -4.34 | 4.05 | .286 |
| | | Valuing math | -20.03 | 19.99 | .318 |
| | | Ready_number*Valuing math | 10.48 | 11.98 | .383 |
| | | Ready_algebra*Valuing math | -3.15 | 7.18 | .661 |
| | | Ready_measurement*Valuing math | 4.95 | 4.55 | .279 |
| | | Ready_geometry*Valuing math | -4.30 | 3.87 | .268 |
| | | Ready_data*Valuing math | 2.04 | 5.94 | .732 |
| | | Homework time | 10.90 | 14.55 | .455 |
| | | Ready_number *Homework time | -33.81 | 14.47 | .021 |
| | | Ready_algebra *Homework time | 8.43 | 6.63 | .206 |
| | | Ready_measurement *Homework time | 3.63 | 9.61 | .706 |
| | | Ready_geometry *Homework time | -0.05 | 6.83 | .994 |
| | | Ready_data *Homework time | 15.46 | 6.61 | .021 |
| | | Home resources | 3.48 | 1.58 | .027 |
| | Random | $\tau_{00}$ | 1935.29 | 43.99 | <.001 |
| | | Self-confidence | 32.35 | 5.69 | .035 |
| | | Valuing math | 100.41 | 10.02 | <.001 |
| | | Homework time | 247.90 | 15.74 | <.001 |
| | | $\sigma^2$ | 1904.28 | 43.64 | |
| 17 | Fixed | INT | 496.57 | 14.45 | <.001 |
| | | Professional development | -3.59 | 3.00 | .233 |
| | | Extra lessons | -15.61 | 1.06 | <.001 |
| | | Self-confidence | 26.61 | 3.58 | <.001 |
| | | Professional development*Self-confidence | -0.08 | 0.85 | .923 |
| | | Valuing math | -5.22 | 5.33 | .330 |
| | | Professional development*Valuing math | 1.25 | 1.21 | .301 |
| | | Homework time | -7.18 | 5.32 | .179 |
| | | Professional development*Homework time | 1.05 | 1.33 | .431 |
| | | Home resources | 3.52 | 1.58 | .026 |
| | Random | $\tau_{00}$ | 1927.98 | 43.91 | <.001 |
| | | Self-confidence | 30.29 | 5.50 | .037 |
| | | Valuing math | 92.78 | 9.63 | <.001 |
| | | Homework time | 253.49 | 15.92 | <.001 |
| | | $\sigma^2$ | 1905.96 | 43.66 | |

When comparing the proportion of variance accounted for by Models 15-17 with that of the foundational level-1 model (Model 9), it appears that Model 16 was the most efficient one (see Table 20). As an example, whereas the inclusion of ready to teach math topics in Model 16 resulted in a reduction of 4% in the between school variance to be explained, the addition of math-related professional development resulted in an increase of 3% of the between school variance to be explained. No improvement in the within school variance was noted by use of these models.

Table 21.
*Comparison of $R^2$ between Level-2 Teacher Background and Foundational Level-1 Model for USA*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 15 vs. 9 | 0.00 | 0.00 |
| 16 vs. 9 | 0.04 | 0.00 |
| 17 vs. 9 | -0.03 | 0.00 |

When including all the teacher-related variables (i.e., preparation to teach, ready to teach math topics, and math-related professional development) in Model 18 to predict math achievement, one statistically significant cross-level interaction effect was produced (see Table 22). Specifically, ready to teach number was found to interact with time student spent on homework ($\gamma$ = -.33.73, *SE* = 15.54, *p* = .031). Also, in this model, all the random effects were statistically significant, meaning that a considerable amount of variance remained to be explained within and between schools.

Table 22.
*Parameter Estimates for the Combined Teacher Background Model for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 18 | Fixed | INT | 472.15 | 47.76 | <.001 |
| | | Preparation | -7.23 | 10.20 | .479 |
| | | Professional development | -3.45 | 2.73 | .209 |
| | | Ready_number | -33.50 | 35.68 | .350 |
| | | Ready_algebra | 3.70 | 22.42 | .870 |
| | | Ready_measurement | -0.40 | 20.33 | .984 |
| | | Ready_geometry | 32.16 | 20.50 | .119 |
| | | Ready_data | 14.32 | 22.43 | .524 |

Table 22.
*Parameter Estimates for the Combined Teacher Background Model for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Extra lessons | -15.62 | 0.94 | <.001 |
| | | Self-confidence | 39.58 | 12.57 | .002 |
| | | Preparation*Self-confidence | -4.10 | 2.72 | .133 |
| | | Professional development*Self-confidence | 0.02 | 0.74 | .980 |
| | | Ready_number *Self-confidence | -10.42 | 9.01 | .250 |
| | | Ready_algebra *Self-confidence | 8.99 | 5.78 | .122 |
| | | Ready_measurement *Self-confidence | -3.20 | 5.19 | .538 |
| | | Ready_geometry *Self-confidence | 2.59 | 5.36 | .630 |
| | | Ready_data *Self-confidence | -3.13 | 5.68 | .582 |
| | | Valuing math | -21.20 | 17.07 | .217 |
| | | Preparation*Valuing math | 5.88 | 3.61 | .105 |
| | | Professional development*Valuing math | 1.05 | 0.96 | .276 |
| | | Ready_number *Valuing math | 10.78 | 12.32 | .383 |
| | | Ready_algebra *Valuing math | -5.95 | 8.13 | .466 |
| | | Ready_measurement *Valuing math | 6.55 | 6.95 | .348 |
| | | Ready_geometry *Valuing math | -5.59 | 7.34 | .448 |
| | | Ready_data *Valuing math | 0.50 | 7.27 | .945 |
| | | Homework time | 8.88 | 20.25 | .661 |
| | | Preparation*Homework time | 3.10 | 4.72 | .513 |
| | | Professional development*Homework time | 0.96 | 1.24 | .442 |
| | | Ready_number *Homework time | -33.73 | 15.54 | .031 |
| | | Ready_algebra *Homework time | 6.84 | 9.89 | .490 |
| | | Ready_measurement *Homework time | 4.19 | 9.21 | .650 |
| | | Ready_geometry *Homework time | -0.84 | 8.89 | .926 |
| | | Ready_data *Homework time | 15.18 | 9.53 | .113 |
| | | Home resources | 3.42 | 1.47 | .020 |
| | Random | $\tau_{00}$ | 1946.84 | 44.12 | <.001 |
| | | Self-confidence | 32.85 | 5.73 | .034 |
| | | Valuing math | 95.69 | 9.78 | <.001 |
| | | Homework time | 250.65 | 15.83 | <.001 |
| | | $\sigma^2$ | 1903.98 | 43.63 | |

As evident in Table 23, Model 18 appeared to be more efficient than Models 9, 15, and 17 in terms of the amount of explained variance accounted for between schools. Specifically, an increase of 2% to 5% in the between school variance was likely to result when using Model 18 as opposed to Models 9, 15, or 17. However, when compared to Model 16, a decrease of 2% in the between school variance was noted in Model 18.

Table 23.

*Comparison of $R^2$ between Model 18 and Previously Constructed Models 9 and 15-17*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 18 vs. 9 | 0.02 | 0.00 |
| 18 vs. 15 | 0.02 | 0.00 |
| 18 vs. 16 | -0.02 | 0.00 |
| 18 vs. 17 | 0.05 | 0.00 |

As shown in Figure 13, the interaction between teacher ready to teach number and time student spent on homework suggested that students' math achievement was inversely related to time student spent on homework. That is, the less time students spent on homework, the better they seemed to perform in math. Interestingly, this pattern of relationship was observed for both groups of students (i.e., those with ready to teach teachers and those with very ready to teach teachers). Unexpectedly, in comparing two groups of students, the one with teachers who reported ready to teach the number topic consistently achieved higher math scores than the other group of students whose teachers reported very ready to teach the subject. The size of differences in math achievement between the two groups, however, was small when students spent a small amount of time on homework and became substantially large when they spent a high amount of time on homework.

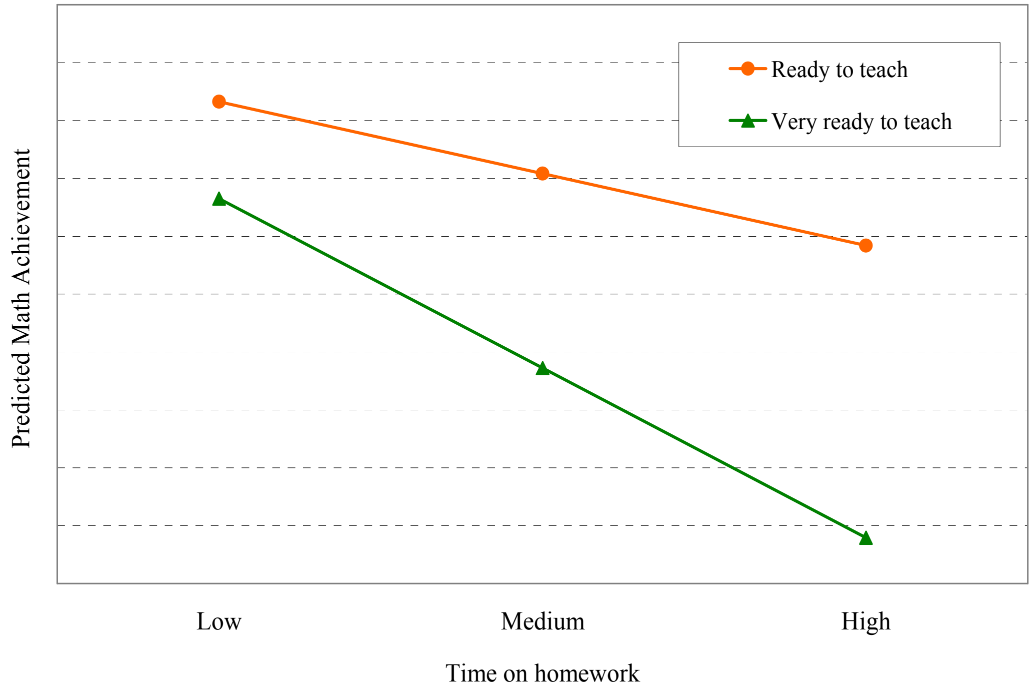



*Figure 13.* Interaction between Time Student Spent on Homework and Teacher Reported Readiness to Teach Number for USA

*Research Question 5*

To what extent are school-related variables (i.e., class size, school resources for math instruction, and teacher perception of math instructional limitations due to student factors) associated with TIMSS 2003 eighth-grade math scores in each country?

Table 24 provides a summary of the results for Models 19-21 where school-related variables (i.e., class size, school resources for math instruction, and teacher perception of math instructional limitations due to student factors) were separately included in the models to predict math achievement. It was found that, of the three variables, teacher perception of math instructional limitations due to student factors in Model 20 was the only one that significantly contributed to the prediction of math achievement ($\gamma$ = -22.77, *SE* = 6.74, *p* = .001). This means the more limitations due to student factors that the teacher perceived to have with teaching math, the poorer the students tended to achieve in math. Specifically, for every unit increase in teacher perception of instructional limitations due to student factors, it was expected that the students would lower their math scores by 22.77 points after controlling for all the level-1 but not other level-2 variables in the model.

Also, it was observed that Model 19 produced two statistically significant cross-level interaction effects, with one between class size for math instruction and student self-confidence in learning math ($\gamma$ = -4.97, *SE* = 1.93, *p* = .011) and the other between class size for math instruction and student valuing of math ($\gamma$ = 5.52, *SE* = 2.64, *p* = .038). In terms of random effects, all were found statistically significant. One exception was the slope variance of self-confidence in learning math in Model 19, which was not statistically significant ($\tau$ = 10.44, *SE* = 3.23, *p* = .116), meaning that the relationship between self-confidence in learning math and math achievement tended to be similar across schools in the U.S.

Table 24.
*Parameter Estimates for School Background Models for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 19 | Fixed | INT | 483.56 | 8.59 | <.001 |
| | | Class size | -2.08 | 9.21 | .822 |
| | | Extra lessons | -15.67 | 1.05 | <.001 |
| | | Self-confidence | 28.96 | 1.50 | <.001 |
| | | Class size*Self-confidence | -4.97 | 1.93 | .011 |
| | | Valuing math | -3.37 | 2.64 | .204 |
| | | Class size*Valuing math | 5.52 | 2.64 | .038 |
| | | Homework time | -2.53 | 2.83 | .375 |
| | | Class size*Homework time | -1.28 | 3.14 | .683 |
| | | Extra lessons | 3.66 | 1.58 | .020 |
| | Random | $\tau_{00}$ | 1959.93 | 44.27 | <.001 |
| | | Self-confidence | 10.44 | 3.23 | .116 |
| | | Valuing math | 87.62 | 9.36 | <.001 |
| | | Homework time | 253.26 | 15.91 | <.001 |
| | | $\sigma^2$ | 1911.06 | 43.72 | |
| 20 | Fixed | INT | 496.39 | 8.28 | <.001 |
| | | Instructional limitation | -22.77 | 6.74 | .001 |

Table 24.
*Parameter Estimates for School Background Models for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Extra lessons | -15.53 | 1.06 | <.001 |
| | | Self-confidence | 25.61 | 1.92 | <.001 |
| | | Instructional limitation*Self-confidence | 1.21 | 1.64 | .463 |
| | | Valuing math | 0.24 | 2.52 | .925 |
| | | Instructional limitation*Valuing math | -1.08 | 2.23 | .627 |
| | | Homework time | -4.70 | 2.89 | .106 |
| | | Instructional limitation*Homework time | 2.56 | 2.76 | .356 |
| | | Home resources | 3.55 | 1.58 | .024 |
| | Random | $\tau_{00}$ | 1683.59 | 41.03 | <.001 |
| | | Self-confidence | 29.28 | 5.41 | .038 |
| | | Valuing math | 95.34 | 9.76 | <.001 |
| | | Homework time | 254.32 | 15.95 | <.001 |
| | | $\sigma^2$ | 1906.00 | 43.66 | |
| 21 | Fixed | INT | 493.60 | 13.88 | <.001 |
| | | School resources | -7.48 | 8.60 | .386 |
| | | Extra lessons | -15.59 | 1.06 | <.001 |
| | | Self-confidence | 25.46 | 3.70 | <.001 |
| | | School resources*Self-confidence | 0.60 | 2.48 | .809 |
| | | Valuing math | -4.28 | 5.51 | .438 |
| | | School resources*Valuing math | 2.64 | 3.21 | .413 |
| | | Homework time | -9.96 | 5.88 | .092 |
| | | School resources*Homework time | 4.57 | 3.82 | .234 |
| | | Home resources | 3.59 | 1.58 | .024 |
| | Random | $\tau_{00}$ | 1930.61 | 43.94 | <.001 |
| | | Self-confidence | 29.73 | 5.45 | .038 |
| | | Valuing math | 91.01 | 9.54 | <.001 |
| | | Homework time | 249.80 | 15.81 | <.001 |
| | | $\sigma^2$ | 1906.71 | 43.67 | |

In comparing Models 19-21 with Model 9 in terms of the proportion of variance accounted for, none of these models worked better than Model 9 (see Table 25). Whereas Model 9 accounted for 1% more of the within variance than Model 19, it accounted for 3% more of the between variance than Model 20, and 2% more of the between variance than Model 21.

Table 25.
*Comparison of $R^2$ between Level-2 Teacher Background and Foundational Level-1 Model for USA*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 19 vs. 9 | 0.00 | -0.01 |
| 20 vs. 9 | -0.03 | 0.00 |
| 21 vs. 9 | -0.02 | 0.00 |

Similar to Model 19, the combined model with all of the school background-related predictors produced two statistically significant cross-level interaction effects. One interaction was between class size for math instruction and student self-confidence in learning math ($\gamma$ = -5.20, *SE* = 1.95, *p* = .009) and the other interaction was between class size for math instruction and student valuing of math ($\gamma$ = 5.23, *SE* = 2.58, *p* = .044). Again, in Model 22, the slope variance of self-confidence in learning math was not statistically significant ($\tau$ = 23.18, *SE* = 4.81, *p* = .104) whereas the remaining random effects were statistically significant. This indicated that the relationship between school background-related variables and math achievement were not statistically significantly different across schools in the U.S.

Table 26.
*Parameter Estimates for the Combined School Background Model for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | Fixed | INT | 515.93 | 14.62 | <.001 |
| | | Instructional limitation | -24.90 | 6.49 | <.001 |
| | | Class size | 2.25 | 8.20 | .784 |
| | | School resources | -13.00 | 8.54 | .130 |
| | | Extra lessons | -15.58 | 1.05 | <.001 |
| | | Self-confidence | 26.24 | 3.74 | <.001 |
| | | Instructional limitation*Self-confidence | 1.67 | 1.60 | .301 |
| | | Class size*Self-confidence | -5.20 | 1.95 | .009 |
| 22 | | School resources*Self-confidence | 1.26 | 2.39 | .599 |
| | | Valuing math | -5.55 | 6.79 | .415 |
| | | Instructional limitation*Valuing math | -0.98 | 2.32 | .671 |
| | | Class size*Valuing math | 5.23 | 2.58 | .044 |
| | | School resources*Valuing math | 2.01 | 3.27 | .539 |
| | | Homework time | -12.08 | 6.77 | .076 |
| | | Instructional limitation*Homework time | 3.18 | 2.84 | .265 |
| | | Class size*Homework time | -2.36 | 3.19 | .461 |
| | | School resources*Homework time | 5.50 | 3.91 | .162 |

Table 26.
*Parameter Estimates for the Combined School Background Model for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Home resources | 3.56 | 1.56 | .022 |
| | Random | $\tau_{00}$ | 1688.59 | 41.09 | <.001 |
| | | Self-confidence | 23.18 | 4.81 | .104 |
| | | Valuing math | 93.15 | 9.65 | <.001 |
| | | Homework time | 255.36 | 15.98 | <.001 |
| | | $\sigma^2$ | 1903.64 | 43.63 | |

As shown in Table 27, Model 22 appears to be less efficient than Models 9 and 19-21. Whereas the amount of explained variance within schools in Model 22 did not change compared to these models (pseudo $R^2 = 0$), the amount of explained variance between schools in Model 22 decreased by 4% (compared to Model 20) to 7% (compared to Models 9 and 19).

Table 27.
*Comparison of $R^2$ between Model 22 and Previously Constructed Models 9 and 19-21 for USA*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 22 vs. 9 | -0.07 | 0.00 |
| 22 vs. 19 | -0.07 | 0.00 |
| 22 vs. 20 | -0.04 | 0.00 |
| 22 vs. 21 | -0.05 | 0.00 |

The modeled means of predicted math achievement for the two interactions observed in Model 22 are displayed in Figures 14-15. The data in Figure 14 suggested that for students who reported having high self-confidence in learning math, changes from small class size (i.e., 1-24 students) to large class size (i.e., 41+ students) tended to lower their math scores significantly. Conversely, for students who reported having low self-confidence in learning math, increases in class size appeared to improve their math scores. Thus, it appears math achievement gap among eighth-grade students with different levels of self-confidence in learning math was most substantial when they learned math in small class size and became smaller when they learned math in large class size.

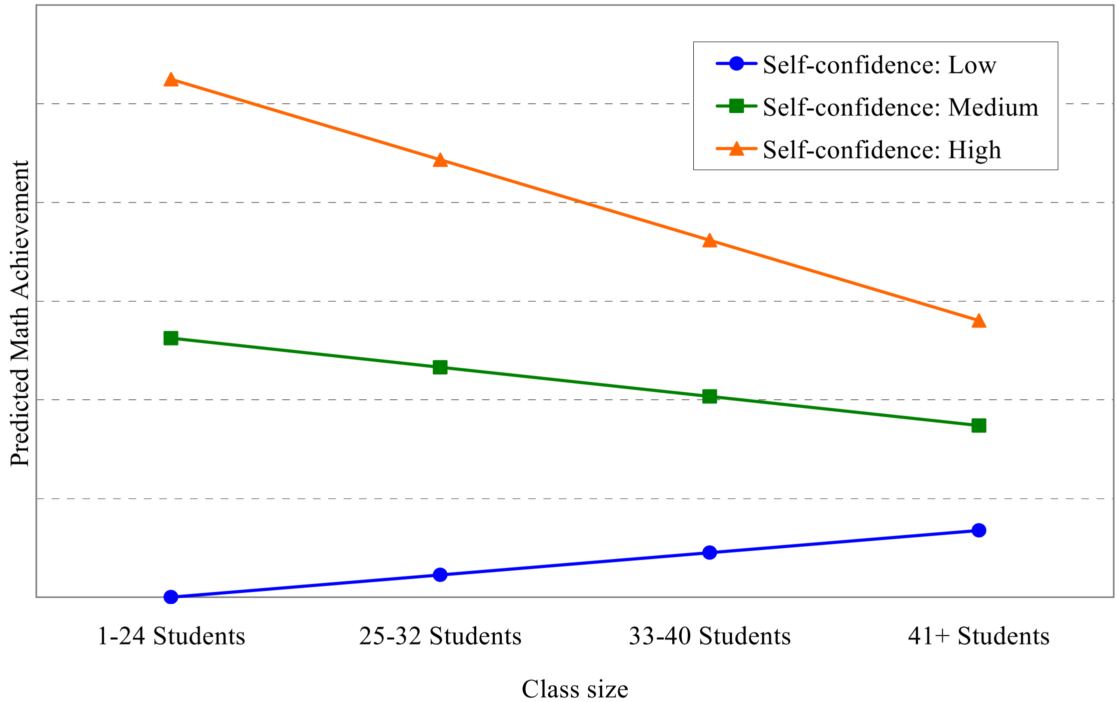



*Figure 14.* Interaction between Class Size for Math Instruction and Self-confidence in Learning Math for USA

Interestingly, as shown in Figure 15, in schools with small class size (i.e., 1-24 students), students who reported having a low level of valuing math tended to achieve higher math scores than their peers who reported having medium or high levels of valuing math. This pattern of relationship, however, appears to reverse in schools with larger classes. That is, students with medium and high levels of valuing math tended to perform better than their peers who reported having a low level of valuing math. Nevertheless, a similar trend was noted for all of the students, regardless of their levels of valuing math. That is, changes from smaller classes to big classes were associated with increased math scores.

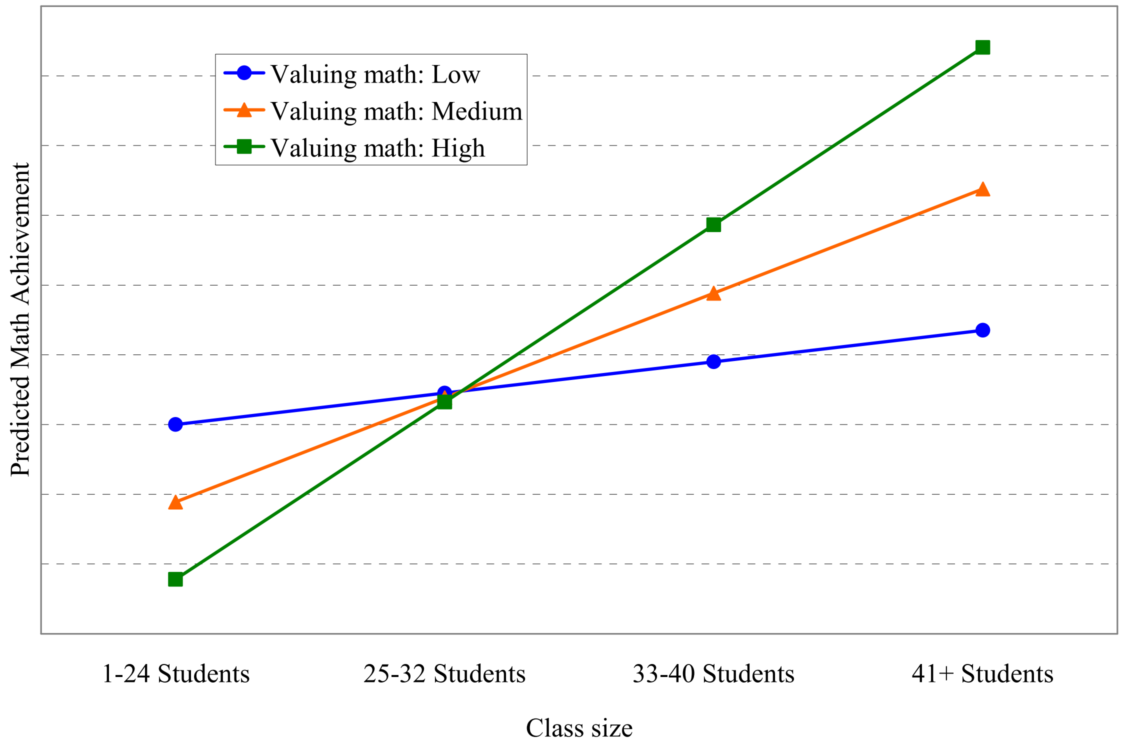



*Figure 15.* Interaction between Class Size for Math Instruction and Valuing of Math for USA

*Final Model*

With an intention to identify the most efficient and parsimonious model to predict eighth-grade math achievement in the USA, Model 23 was built and compared with the three combined models, Models 14, 18, and 22. It is also worth noting that in Model 23, the slope variance for student self-confidence in learning math was set to 0 because it was not statistically significant in Models 14 and 22.

As can be seen from Table 28, Model 23 produced several statistically significant fixed effects. The three significant level-2 main effects included teacher perception of math instructional limitations due to student factors ($\gamma$ = 32.35, *SE* = 7.29, *p* <.001), opportunity to learn data ($\gamma$ = -.53, *SE* = .22, *p* = .015), and opportunity to learn geometry ($\gamma$ = .37, *SE* = .16, *p* = .025). The three level-1 significant main effects were:

extra math lessons ($\gamma$ = -15.43, *SE* = 1.06, *p* <.001), self-confidence in learning math ($\gamma$ = 26.63, *SE* = 1.26, *p* <.001), and home resources ($\gamma$ = 3.65, *SE* = 1.54, *p* = .018). And, the two significant cross-level interactions were found between opportunity to learn geometry and time student spent on homework ($\gamma$ = -.21, *SE* = .07, *p* = .005) and between teacher ready to teach number and time student spent on homework ($\gamma$ = -38.58, *SE* = 18.95, *p* = .043). Also, in this model, all the random effects were statistically significant.

Table 28.
*Parameter Estimates for Full Model for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 23 | Fixed | INT | 393.91 | 167.34 | .020 |
| | | Homework assignment | 32.35 | 7.29 | <.001 |
| | | Instructional limitation | -11.75 | 6.23 | .061 |
| | | Class size | -2.81 | 6.74 | .677 |
| | | Opportunity_algebra | 0.43 | 0.22 | .050 |
| | | Opportunity_data | -0.53 | 0.22 | .015 |
| | | Opportunity_geometry | 0.37 | 0.16 | .025 |
| | | Opportunity_measurement | -0.05 | 0.22 | .829 |
| | | Opportunity_number | 0.47 | 1.86 | .800 |
| | | Ready_ number | -19.05 | 30.45 | .532 |
| | | Ready_ algebra | -7.44 | 14.37 | .605 |
| | | Ready_ measurement | 6.46 | 19.89 | .746 |
| | | Ready_ geometry | 13.68 | 20.96 | .515 |
| | | Ready_ data | 6.48 | 18.44 | .726 |
| | | Extra lessons | -15.43 | 1.06 | <.001 |
| | | Self-confidence | 26.63 | 1.26 | <.001 |
| | | Valuing math | -3.80 | 46.33 | .935 |
| | | Homework assignment*Valuing math | -3.15 | 2.60 | .228 |
| | | Instructional limitation*Valuing math | -1.16 | 2.24 | .605 |
| | | Class size*Valuing math | 3.75 | 2.54 | .143 |
| | | Opportunity_algebra *Valuing math | 0.02 | 0.09 | .777 |
| | | Opportunity_ data *Valuing math | -0.03 | 0.10 | .746 |
| | | Opportunity_ geometry *Valuing math | -0.05 | 0.08 | .554 |
| | | Opportunity_ measurement *Valuing math | 0.14 | 0.11 | .195 |
| | | Opportunity_ number *Valuing math | -0.14 | 0.54 | .799 |
| | | Ready_ number *Valuing math | 9.28 | 10.23 | .366 |
| | | Ready_ algebra | -0.92 | 7.71 | .905 |
| | | Ready_measurement | 2.22 | 6.04 | .714 |
| | | Ready_ geometry | -1.71 | 4.47 | .703 |
| | | Ready_ data | -2.89 | 6.42 | .653 |

Table 28.
*Parameter Estimates for Full Model for USA*

| Model | Effect | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Homework time | -14.22 | 60.67 | .815 |
| | | Homework assignment*Homework time | -3.44 | 3.53 | .332 |
| | | Instructional limitation*Homework time | 5.69 | 3.16 | .073 |
| | | Class size*Homework time | -5.09 | 3.42 | .139 |
| | | Opportunity_algebra*Homework time | 0.03 | 0.13 | .798 |
| | | Opportunity_data*Homework time | 0.17 | 0.13 | .198 |
| | | Opportunity_geometry*Homework time | -0.21 | 0.07 | .005 |
| | | Opportunity_measurement*Homework time | 0.20 | 0.13 | .127 |
| | | Opportunity_number*Homework time | 0.29 | 0.75 | .696 |
| | | Ready_number*Homework time | -38.58 | 18.95 | .043 |
| | | Ready_ algebra | 10.40 | 8.06 | .199 |
| | | Ready_measurement | -3.00 | 10.58 | .777 |
| | | Ready_geometry | 2.52 | 8.49 | .767 |
| | | Ready_data | 12.29 | 7.65 | .110 |
| | | Home resources | 3.65 | 1.54 | .018 |
| | Random | $\tau_{00}$ | 1371.00 | 37.03 | <.001 |
| | | Valuing math | 99.05 | 9.95 | <.001 |
| | | Homework time | 225.58 | 15.02 | <.001 |
| | | $\sigma^2$ | 1918.98 | 43.81 | |

Surprisingly, when comparing this full model with earlier combined models, it was found that Model 23 was more efficient than Model 18 and 22 but less efficient than Model 14. Specifically, whereas the amount of between school variance accounted for by Model 23 was reduced by 17% as compared to Model 18 and 24% as compared to Model 22, it increased by 8% as compared to Model 14. Therefore, for USA, Model 14 serves as the best model for predicting math achievement.

Table 29.
*Comparison of $R^2$ between Model 23 and Previously Constructed Models 14, 18 and 22 For USA*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 14 | -0.08 | -0.01 |
| 18 | 0.17 | -0.01 |
| 22 | 0.24 | -0.01 |

Figure 16 visually displays the nature of the cross-level interactions between opportunity to learn geometry and time student spent on homework. Interestingly, in the presence of other level-2 factor predictors in the model (i.e., student background, student home resources, instructional practice, teacher background, and school-background), this interaction had a different pattern from what was observed in Model 14 where only instructional practice-related predictors were included in the model. Specifically, Model 23 suggested that when the opportunity to learn geometry was low students tended to score low in math, and that the achievement gaps across students with different levels of time on homework were relatively small. However, as the opportunity to learn increased, students who spent a medium or low amount of time on homework tended to gain higher math scores whereas students who spent a high amount of time on homework tended to perform slightly poorer in math.

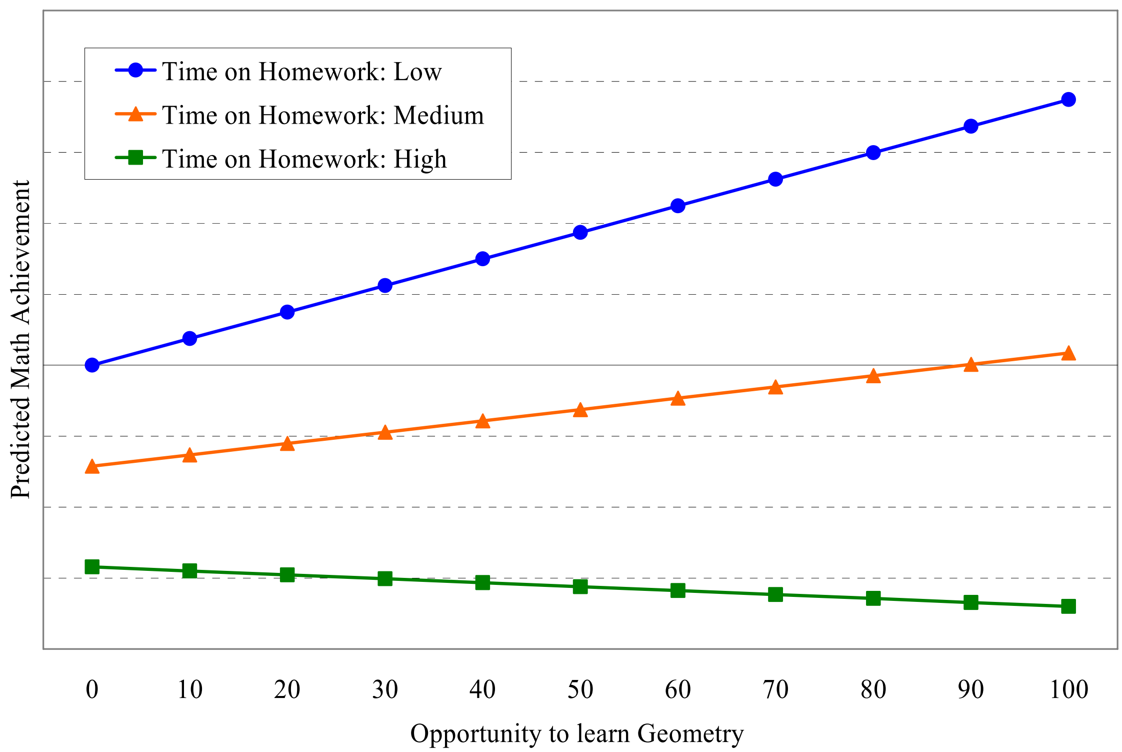



*Figure 16.* Interaction between Opportunity to Learn Geometry by Time Student Spent on Homework for USA

Figure 17 presents the model predicted math achievement based upon teacher ready to teach number and time student spent on homework. Unexpectedly, across the amounts of time on homework, students whose teachers reported very ready to teach number tended to achieve lower math scores than their peers whose teachers reported ready to teach number. The achievement gap between these students, however, was small when their time on homework was low and became more substantial when the amount of homework was high. It is important to note that this graph was constructed based on 149 teachers who reported very ready to teach number and only 4 teachers who reported ready to teach number.

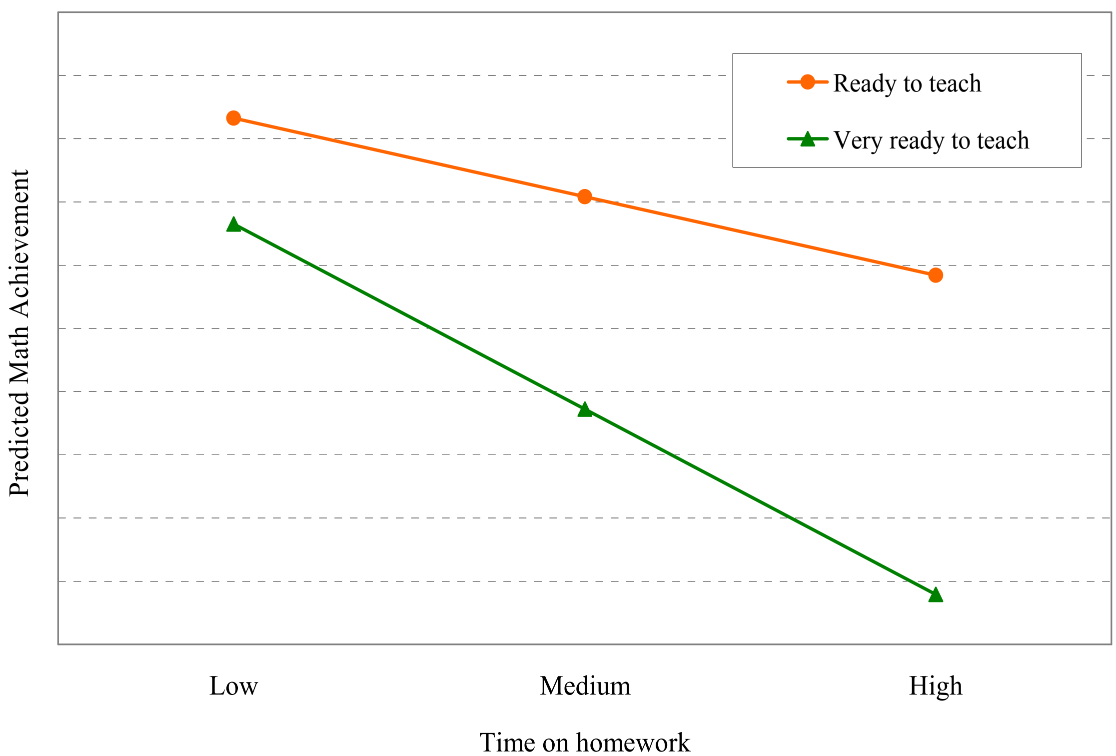



*Figure 17.* Interaction between Teacher Reported Ready to Teach Number by Time Student Spent on Homework for USA

## Results for Canada

### *Evaluation of Missing Data*

As a result of the listwise deletion method, the sample size for Canada was reduced from 8,473 students and 354 schools to 6,248 students and 271 schools. This means approximately 73.74% of the original sample had complete data on all variables of interest in this study. In order to evaluate the extent to which the data for Canada were missing completely at random, the missingness on 19 level-2 variables was correlated. Results of this analysis suggested a non-randomness of missing data, with the majority of the correlation coefficients larger than .70, indicating strong positive relationships among missingness indicators of the variables. In addition, when missingness was correlated with values of itself as well as values of other variables, only weak correlations were observed ($r$ = -.15 to .23, $p$ = .005). In summary, the missing data mechanism for Canada was not missing completely at random.

### *Univariate Analysis*

A descriptive examination of level-1 variables (i.e., overall math achievement, gender, self-confidence in learning math, valuing of math, time on math homework, extra math lessons, and home resources for learning math) was conducted using SAS 9.13. Of the complete sample of 6,248 eighth-grade students, 3,092 (49.49%) were male and 3,156 (50.51%) were female. On average, the weighted overall math achievement for Canadian students was 529.30 ($SD$ = 61.17) with the lowest score of 322.82 and the highest score of 728.39 (see Table 30).

With regard to level-1 predictor variables, it appeared that, on average, eighth-grade students in Canada had good support at home in terms of resources for learning (*M* = 2.85 , *SD* = .41), were above medium level of self-confidence in learning math (*M* = 1.47, *SD* = .75) and valuing of math (*M* = 1.58, *SD* = .61), spent little time on math homework (*M* = .85, *SD* = .58), and only had extra math lessons occasionally (*M* = .45, *SD* = .76) (see Table 30).

Table 30.

*Weighted Descriptive Statistics for Level-1 Variables for Canada (N = 6,248)*

| Variable | *M* | *SD* | Min | Max |
|---|---|---|---|---|
| Overall math achievement | 529.30 | 61.18 | 322.82 | 728.39 |
| Self-confidence in learning math | 1.47 | 0.75 | 0 | 2 |
| Valuing of math | 1.58 | 0.61 | 0 | 2 |
| Time on math homework | 0.85 | 0.58 | 0 | 2 |
| Extra math lessons | 0.45 | 0.76 | 0 | 3 |
| Home resources for learning math | 2.85 | 0.41 | 0 | 3 |

Note: When weight was used to compute means in SAS, skewness and kurtosis were not produced

In terms of distributions of level-1 variables, the unweighted descriptive results from Table 31 suggested that all but two variables, extra math lessons and home resources for learning math, approximated normality, with skewness and kurtosis values within the range of -1.00 and 1.00.

Table 31.

*Unweighted Descriptive Statistics for Level-1 Variables for Canada (N = 6,248)*

| Variable | *M* | *SD* | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| Overall math achievement | 528.34 | 60.91 | 322.82 | 728.39 | -0.07 | -0.30 |
| Self-confidence in learning math | 1.44 | 0.75 | 0 | 2 | -0.92 | -0.63 |
| Valuing of math | 1.56 | 0.61 | 0 | 2 | -1.05 | 0.08 |
| Time on math homework | 0.84 | 0.60 | 0 | 2 | 0.08 | -0.38 |
| Extra math lessons | 0.44 | 0.77 | 0 | 3 | 1.71 | 2.13 |
| Home resources for learning math | 2.85 | 0.42 | 0 | 3 | -3.03 | 10.37 |

Similarly, a descriptive analysis was conducted on the 19 predictor variables at the school level. As shown in Table 32, on average, Canadian students had a moderate to high percentage of opportunity to learn math content domains, ranging from 59.14 (*SD* = 28.26) for algebra to 95.85 (*SD* = 8.88) for number. Although less than half of math teachers reported being prepared to teach math content (*M* = .44, *SD* = .50), on average, they participated in various types of math-related professional development (*M* = 2.87, *SD* =1.81) and reported a high level of readiness to teach (*M* = 1.75, *SD* = .44 for data to *M* = 1.94, *SD* = .24 for number).

The data also suggested that in nearly half of the lessons, students were given activities related to math instructional practice (*M* = 1.81, *SD* = .27) and math content (*M* = 1.70, *SD* = .22). On average, a medium amount of homework was assigned to the students (*M* = 1.14, *SD* = .61). Finally, class size in Canadian schools tended to be small, less than 33 students (*M* = .73, *SD* = .62) and teachers' perception of instructional limitations due to student factors was low (*M* = .52, *SD* = .70). On average, the availability of school resources for math instruction was relatively high (*M* = 1.41, *SD* = .56). Noticeably, across 271 schools, the average math instructional hours per year varied greatly, ranging from 30 to 388, with a mean of 161.57 (*SD* = 44.72).

Also, as shown in Table 32, 12 out of the 19 level-2 predictor variables appeared to have approximately normal distributions with skewness and kurtosis values within the normality approximation range of -1.00 to 1.00. The seven variables that appeared to depart from normality included opportunity to learn number, preparation to teach, ready to teach number, algebra, measurement, geometry, and data, and average math instructional hours per year.

Table 32.

*Unweighted Descriptive Statistics for Level-2 Variables for Canada (N = 271)*

| Variable | *M* | *SD* | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| Opportunity to learn number | 95.85 | 8.88 | 40 | 100 | -3.01 | 11.41 |
| Opportunity to learn algebra | 59.14 | 28.26 | 0 | 100 | -0.18 | -0.75 |
| Opportunity to learn measurement | 77.04 | 23.23 | 0 | 100 | -1.00 | 0.36 |
| Opportunity to learn geometry | 72.86 | 19.12 | 0 | 100 | -0.92 | 1.24 |
| Opportunity to learn data | 60.61 | 32.30 | 0 | 100 | -0.35 | -1.15 |
| Amount of homework assignment | 1.14 | 0.61 | 0 | 2 | -0.09 | -0.42 |
| Instructional practice-related activities in math lessons | 1.81 | 0.27 | 1.13 | 2.71 | 0.18 | 0.01 |
| Content-related activities in math lessons | 1.70 | 0.22 | 1.04 | 2.32 | 0.21 | 0.43 |
| Preparation to teach | 0.44 | 0.50 | 0 | 1 | 0.26 | -1.95 |
| Ready to teach number | 1.94 | 0.24 | 1 | 2 | -3.76 | 12.25 |
| Ready to teach algebra | 1.82 | 0.39 | 0 | 2 | -1.89 | 2.15 |
| Ready to teach measurement | 1.83 | 0.39 | 0 | 2 | -1.92 | 2.32 |
| Ready to teach geometry | 1.88 | 0.35 | 0 | 2 | -2.84 | 7.67 |
| Ready to teach data | 1.75 | 0.44 | 0 | 2 | -1.26 | -0.05 |
| Math-related professional development | 2.87 | 1.81 | 0 | 5 | -0.35 | -1.25 |
| Class size for math instruction | 0.73 | 0.62 | 0 | 2 | 0.26 | -0.62 |
| School resources for math instruction | 1.41 | 0.56 | 0 | 2 | -0.28 | -0.84 |
| Teacher perception of math instructional limitations due to student factors | 0.52 | 0.70 | 0 | 2 | 0.99 | -0.32 |
| Average math instructional hours per year | 161.57 | 44.72 | 30 | 388 | 1.10 | 4.92 |

### *Bivariate Analysis*

The results of weighted bivariate correlations among six level-1 variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, extra math lessons, and home resources for learning math) are presented in Appendix G. It appeared from these results that level-1 predictor variables were uncorrelated from each other, with correlation coefficients ranging from -.24 between self-confidence in learning math and extra math lessons to .37 between self-confidence in learning math and valuing of math. It was interesting to note that gender tended to have a negative albeit weak relationship with all level-1 variables except for home resources for learning math ($r$ = .03).

At level-2, unweighted bivariate relationships were estimated for the 19 predictor variables. The correlation matrix for these variables can be found in Appendix H. Unlike level-1, correlation coefficients of level-2 variables had a wider range, from -.27 between percentage of opportunity to learn measurement and preparation to teach to .53 between ready to teach algebra and ready to teach measurement. As expected, correlation coefficients among the variables measuring the same construct tended to be stronger than those measuring different construct. For example, the correlations ranged from .39 to .53 for ready to teach variables. Interestingly, opportunity to learn number had very weak correlations with other opportunity to learn variables ($r$ = -.08 to .14). Another interesting relationship was observed between the number of math instructional hours per year and other variables where most correlation coefficients were close to 0. Similarly, teachers' perception of math instructional limitations due to student factors was found to have very weak to almost no relationship with other variables ($r$ = -.11 to .09).

*Evaluation of HLM Assumptions*

In order to ensure tenability of results produced by multilevel models in this study, an evaluation of HLM assumptions through visual analysis of both level-1 and level-2 random effects of Model 23 was performed. Model 23 was selected because the results of HLM analysis suggested that it was the most efficient model to predict math achievement in Canada (see HLM Analysis for Canada).

The data from Figure 18 suggested that level-1 residuals approximated a normal distribution. In terms of variance, the scatter plot between level-1 residuals and predicted math achievement, as illustrated in Figure 19, suggested that there was evidence of homogeneity of level-1 variance.

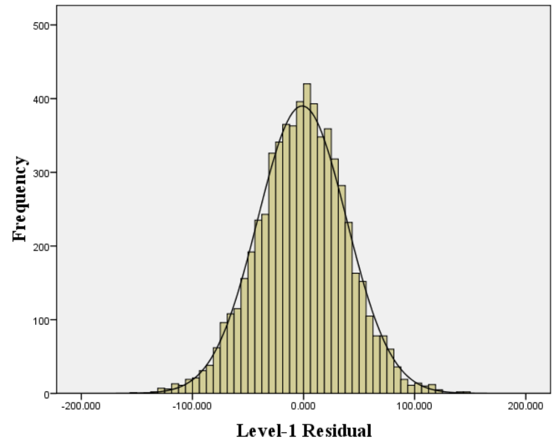



*Figure 18.* Histogram for Level-1 Residuals for Canada

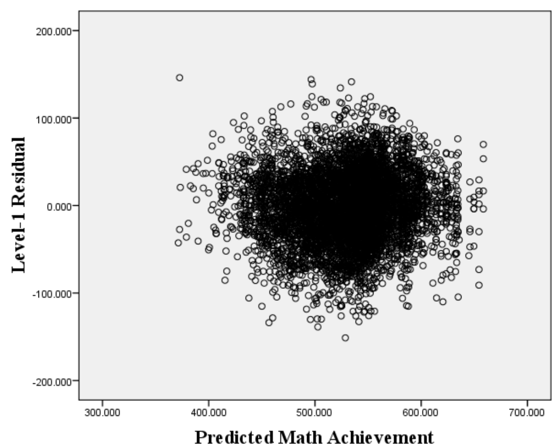



*Figure 19.* Level-1 Residuals by Predicted Math Achievement for Canada

For level-2 random effects, the empirical Bayes residuals for the intercepts and slopes as well as empirical Bayes predicted math scores were used to construct the graphs in Figures 20-27. As can be seen from Figures 20-27, level-2 intercept residuals appeared to have a normal distribution and homogeneous variance.

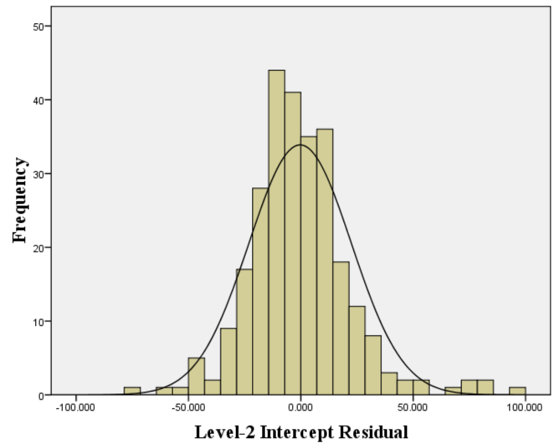



*Figure 20.* Histogram for Level-2 Intercept Residuals for Canada

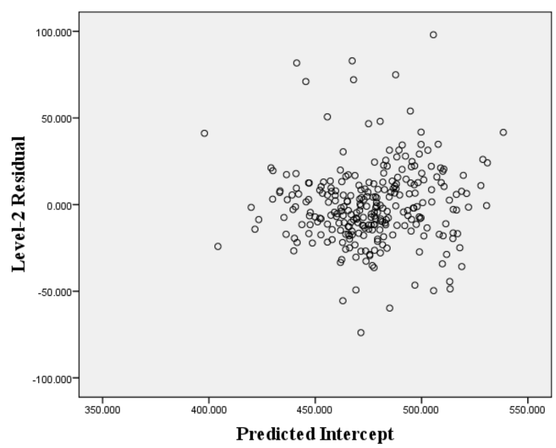



*Figure 21.* Level-2 Intercept Residuals by Predicted Intercept for Canada

Similarly, Figure 22 suggests that level-2 residuals for the slope of Gender approximated a normal distribution and Figure 23 provides evidence of homogeneity of variance.

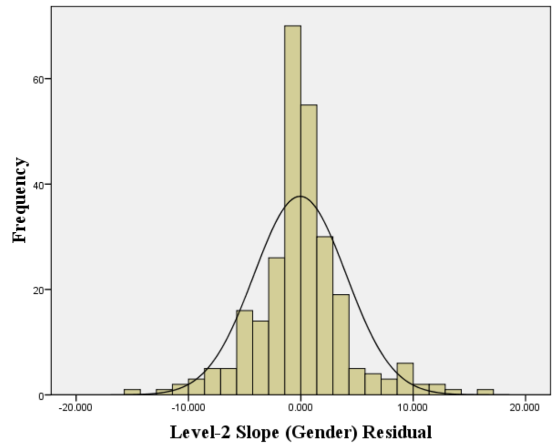



*Figure 22.* Histogram for Level-2 Slope (Gender) Residuals for Canada

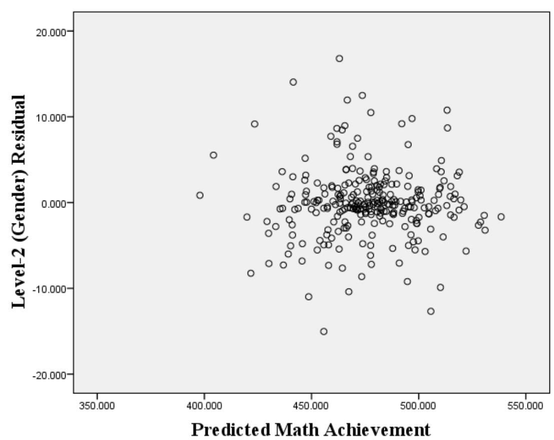



*Figure 23.* Level-2 Slope (Gender) Residuals by Predicted Math Achievement for Canada

Likewise, for the slope of extra math lessons, it can be seen from Figures 24-25 that the slope residuals approximated a normal distribution and had homogeneous variance.

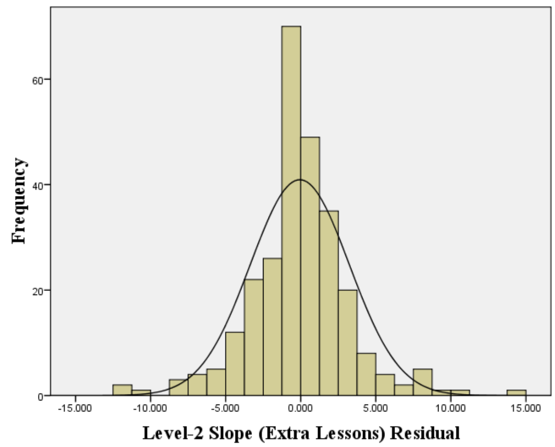



*Figure 24.* Histogram for Level-2 Slope (Extra Lessons) Residuals for Canada

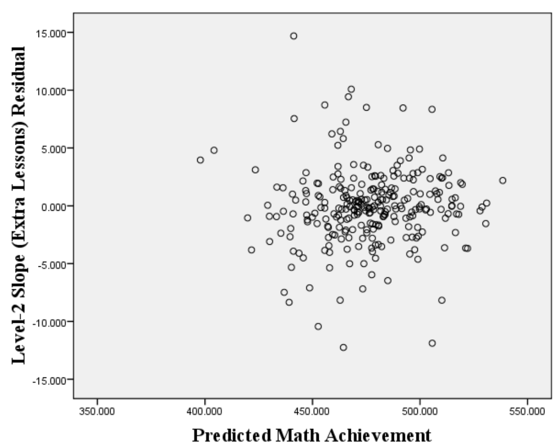



*Figure 25.* Level-2 Slope (Extra Lessons) Residuals by Predicted Math Achievement for Canada

Finally, an examination of Figures 26-27 also suggested that the level-2 residuals for the slope of self-confidence had an approximately normal distribution and their variances across school were relatively homogeneous.

In summary, visual analyses of both level-1 and level-2 random effects suggested that the assumptions of normality and homogeneity of level-1 and level-2 random effects were satisfied.

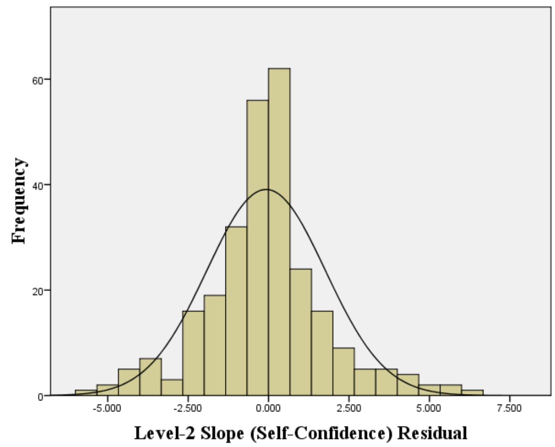



*Figure 26.* Histogram for Level-2 Slope (Self-Confidence) Residuals for Canada

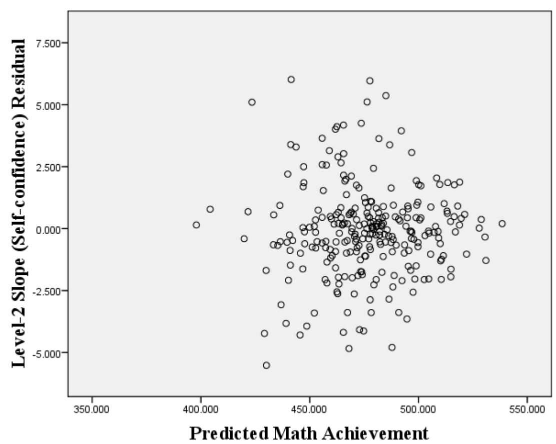



*Figure 27.* Level-2 Slope (Self-Confidence) Residuals by Predicted Math Achievement for Canada

## *HLM Analysis*

### *Unconditional model (Model 1)*

The HLM analysis started with the unconditional model where none of the level-1 or level-2 predictor was included in the model. The results of the unconditional model are presented in Table 33. For Canada, the fixed effect for the intercept was 527.33 (*SE* = 2.55, $p$ <.001). The amount of variability in math achievement was significantly different across schools in Canada ( $\tau_{00}$ = 1,028.87, *SE* = 32.08, $p$ <.001). Within schools, the amount of unexplained variance was much larger than that between schools ($\sigma^2$ = 2,650.40, *SE* = 51.48). The computed intra-class correlation (ICC) of .28 indicated a modest level of natural clustering of students occurred between schools in Canada. In other words, approximately 28% of the total variance in math scores occurred between schools.

Table 33.
*Parameter Estimates for Unconditional Model for Canada*

| Model | Effect | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 1 | Fixed | ICC | 0.28 | | |
| | | INT | 527.33 | 2.55 | <.001 |
| | Random | $\tau_{00}$ | 1028.87 | 32.08 | <.001 |
| | | $\sigma^2$ | 2650.40 | 51.48 | |

*Research Question 1*

To what extent are student background variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, and tutoring in math) associated with TIMSS 2003 eighth-grade math scores in each country?

In order to answer this research question, first, each of the student background variables was entered separately into Model 1 to predict math achievement. Then, as a group of variables, those that contributed significantly in Models 2-6 were included in Model 7 to predict math achievement. Finally, in order to evaluate model fit in terms of proportion of variance accounted for, pseudo $R^2$ was computed for the current model against previously constructed models. Results of these models (Models 2-6) are presented in Table 34.

The data from Table 34 suggested that all of the fixed effects estimated by Models 2-6 were statistically significant, except for that of time on homework in Model 5 ($\gamma$ = 1.98, *SE* = 1.64, *p* = .23). Likewise, all of the random effects estimated in Models 2-6 were statistically significant, except for those of valuing math in Model 4 ($\tau$ = 14.64, *SE* = 3.83, *p* = .44) and time on homework in Model 5 ($\tau$ = 13.85, *SE* = 3.72, *p* = .10). Interestingly, whereas self-confidence in learning math (Model 3) and valuing of math (Model 4) appeared to have positive relationships with math achievement ($\gamma$ = 40.04 and 21.13, *SE* = 1.05 and 1.41; *p* <.001 and .001, respectively), gender (Model 2) and extra math lessons (Model 6) appeared to have negative relationships with math achievement ($\gamma$ = -4.74 and -21.70, *SE* = 2.26 and 1.41, *p* = .036 and <.001, respectively).

An examination of the pseudo $R^2$ across the five models (Models 2-6) suggested that the addition of individual predictors separately to the unconditional model (Model 1) to predict math achievement resulted in a reduction between 0% (Model 5) to 32% (Model 3) for the within school variance. For the between school variance, however, the amount of reduction was smaller, up to 11% (Model 6). In fact, in Models 2-4, the amount of between school variance even increased (i.e., 6% in Model 3 to 15% in Model 2).

Table 34.

*Parameter Estimates for Models 2-6 (Level-1 Student Background) for Canada*

| Model | Effect | Parameters | Estimates | *SE* | *p* | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|
| 2 | Fixed | INT | 529.78 | 2.86 | <.001 | | |
| | | Gender | -4.74 | 2.26 | .036 | | |
| | Random | $\tau_{00}$ | 1186.81 | 34.45 | <.001 | | |
| | | Gender | 195.52 | 13.98 | <.001 | | |
| | | $\sigma^2$ | 2597.50 | 50.97 | | | |
| | Pseudo $R^2$ | | | | | -0.15 | 0.02 |
| 3 | Fixed | INT | 468.80 | 2.83 | <.001 | | |
| | | Self-confidence | 40.04 | 1.05 | <.001 | | |
| | Random | $\tau_{00}$ | 1086.53 | 32.96 | <.001 | | |
| | | Self-confidence | 43.98 | 6.63 | .003 | | |
| | | $\sigma^2$ | 1792.93 | 42.34 | | | |

Table 34.

*Parameter Estimates for Models 2-6 (Level-1 Student Background) for Canada*

| Model | Effect | Parameters | Estimates | *SE* | *p* | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|
| | Pseudo $R^2$ | | | | | -0.06 | 0.32 |
| 4 | Fixed | INT | 493.85 | 3.33 | <.001 | | |
| | | Valuing math | 21.13 | 1.41 | <.001 | | |
| | Random | $\tau_{00}$ | 1117.09 | 33.42 | <.001 | | |
| | | Valuing math | 14.64 | 3.83 | .443 | | |
| | | $\sigma^2$ | 2494.37 | 49.94 | | | |
| | Pseudo $R^2$ | | | | | -0.09 | 0.06 |
| 5 | Fixed | INT | 525.69 | 2.83 | <.001 | | |
| | | Homework time | 1.98 | 1.64 | .229 | | |
| | Random | $\tau_{00}$ | 955.34 | 30.91 | <.001 | | |
| | | Homework time | 13.85 | 3.72 | .096 | | |
| | | $\sigma^2$ | 2645.16 | 51.43 | | | |
| | Pseudo $R^2$ | | | | | 0.07 | 0.00 |
| 6 | Fixed | INT | 536.79 | 2.41 | <.001 | | |
| | | Extra lessons | -21.70 | 1.41 | <.001 | | |
| | Random | $\tau_{00}$ | 914.89 | 30.25 | <.001 | | |
| | | Extra lessons | 79.82 | 8.93 | <.001 | | |
| | | $\sigma^2$ | 2377.11 | 48.76 | | | |
| | Pseudo $R^2$ | | | | | 0.11 | 0.10 |

Note: Pseudo $R^2$ refers to the difference in proportion of variance accounted for between the current models (Models 2-6) and the unconditional model (Model 1).

As a next step of model building, all of the student background variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, and tutoring in math) were included in the combined model, Model 7, to predict math achievement. Interestingly, in the presence of other variables in the model, only three out of five predictors had statistically significant fixed effects. With fixed effect of -12.80 (*SE* = 1.17, $p$ <.001) for extra math lesson, it could be inferred that for each unit increase in extra math lesson (i.e., from 0 for never to 3 for daily), the students were expected to reduce 12.80 points in their math scores while controlling for other predictors in the model. Similarly, with fixed effect of 35.88 (*SE* = 1.19, $p$ <.001) for self-confidence in learning math, it could be interpreted that for each unit increase in level of self-confidence in learning math (i.e., from 0 for low to 2 for high), it was expected that the students would improve 35.88 points in their math scores while controlling for other predictors in the model. Likewise, with fixed effect of 4.81 (*SE* = 1.30, $p$ <.001) for student valuing of math, it could be inferred that for each unit increase in level of student valuing of math (i.e., from 0 for low to 2 for high), it was expected that the students would gain 4.81 more points in their math scores after adjusting for other predictors in the model.

In terms of random effects, all were found statistically significant, except for those of student valuing of math ($\tau$ = 16.81, *SE* = 4.10, $p$ = .50) and time student spent on homework ($\tau$ = 7.38, *SE* = 2.72, $p$ =.50). With the variance for the intercept of 1,381.93 (*SE* = 37.17, $p$ <.001), it could be inferred that statistically significant differences existed across the school means of math achievement after adjusting for the five student background variables in the model. Similarly, it could be interpreted that schools varied significantly in the relationships between math achievement and student gender ($\tau$ = 77.76, *SE* = 8.82, $p$ = .001), extra math lessons ($\tau$ = 48.11, *SE* = 6.94, $p$ <.001), and student self-confidence in learning math ($\tau$ = 53.04, *SE* = 7.28, $p$ <.001).

Table 35.

*Parameter Estimates for Model 7 (Level-1 Student Background) for Canada*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 7 | Fixed | INT | 472.20 | 3.72 | <.001 |
| | | Gender | 1.35 | 1.65 | .415 |
| | | Extra lessons | -12.80 | 1.17 | <.001 |
| | | Self-confidence | 35.88 | 1.19 | <.001 |
| | | Valuing math | 4.81 | 1.30 | <.001 |

Table 35.

*Parameter Estimates for Model 7 (Level-1 Student Background) for Canada*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Homework time | -0.08 | 1.21 | .950 |
| | Random | $\tau_{00}$ | 1381.93 | 37.17 | <.001 |
| | | Gender | 77.76 | 8.82 | .001 |
| | | Extra lessons | 48.11 | 6.94 | <.001 |
| | | Self-confidence | 53.04 | 7.28 | .001 |
| | | Valuing math | 16.81 | 4.10 | >.500 |
| | | Homework time | 7.38 | 2.72 | >.500 |
| | | $\sigma^2$ | 1656.53 | 40.70 | |

An evaluation of model fit was also conducted between Model 7 and previously constructed models, Models 2-6. As expected, the inclusion of student background variables in Model 7 yielded a considerable reduction in amount of variance accounted for in math achievement within schools, from 8% to 37% (see Table 36). Unexpectedly, between schools, the amount of variance appeared to increase notably, from 16% to 51%. In sum, Model 7 was more efficient than earlier models in that it accounted for more variance in math achievement within schools in Canada. However, Model 7 appeared to be less efficient than previously constructed models in that it accounted for less variance in math achievement between schools in Canada.

Table 36.

*Comparison of $R^2$ between Model 7 and Previously Constructed Models for Canada*

| Previous Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 2 | -0.16 | 0.36 |
| 3 | -0.27 | 0.08 |
| 4 | -0.24 | 0.34 |
| 5 | -0.45 | 0.37 |
| 6 | -0.51 | 0.30 |

*Research Question 2*

To what extent are home resources variables (i.e., availability of calculator, computer, and desk for student use) associated with TIMSS 2003 eighth-grade math scores in each country?

Interestingly, when level-1 predictor home resources was added to the unconditional model to predict math achievement, there was a considerable increase of 56% in the between school variance, whereas the amount of reduction in the within school variance was trivial, .4% (see Table 37). In this model, home resources had a statistically significant relationship with math achievement ($\gamma$ = 7.48, *SE* = 1.76, *p* <.001). This means that for every unit increase in home resources (i.e., from 0 to 3), math achievement was expected to increase by 7.48 points, while not controlling for other variables. In addition, with the random effect for home resources being not statistically significant ($\tau$ = 11.51, *SE* = 3.39, *p* > .50), it could be inferred that the relationship between home resources and math achievement tended to be similar across schools in Canada.

Table 37.

*Parameter Estimates for Level-1 Home Resources Model for Canada*

| Model | Effect | Parameters | Estimates | *SE* | *p* | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|
| 8 | Fixed | INT | 506.17 | 5.89 | <.001 | | |
| | | Home resources | 7.48 | 1.76 | <.001 | | |
| | Random | $\tau_{00}$ | 1604.39 | 40.05 | .110 | | |
| | | Home resources | 11.51 | 3.39 | >.500 | | |
| | | $\sigma^2$ | 2640.68 | 51.39 | | | |
| | Pseudo $R^2$ | | | | | -0.56 | 0.004 |

Note: Pseudo $R^2$ refers to the difference in the proportion variance between Model 8 and Model 1.

Given the findings obtained from Models 7 and 8, five out of six student-related variables were entered into the unconditional model to make Model 9. Time student spent on homework was excluded from Model 9 because both of its fixed and random effects were not statistically significant in Model 7. Also, in Model 9, the slope variances of student valuing of math and home resources were set to 0 because they were not statistically significant in earlier models.

As can be seen from Table 38, with the presence of other predictors in Model 9, only extra math lessons, self-confidence in learning math, and student valuing of math had statistically significant relationships with math achievement. Specifically, whereas self-confidence in learning math ($\gamma$ = 35.79, *SE* = 1.20, *p* <.001) and student valuing of math ($\gamma$ = 4.51, *SE* = 1.28, *p* <.001) were positively related to math achievement, an inverse relationship was observed between math achievement and extra math lessons ($\gamma$ = -12.89, *SE* = 1.17, *p* <.001). This could be interpreted to mean that the more self-confidence students expressed in learning math and the higher value students placed in math, the better they achieved in math performance. However, it appears that the more frequently students took extra math lessons, the poorer math scores they achieved.

In terms of random effects, all were found statistically significant, suggesting that the relationships between level-1 predictors (i.e., gender, extra math lessons, and self-confidence) and math achievement varied significantly across schools in Canada.

As compared to Model 7, Model 9 appeared more efficient in that it accounted for more variance between schools (11%), even though a marginal increase in the variance within schools (1%) was noted. Compared to Model 8, Model 9 accounted for a significantly higher amount of the variance within school (37%) and the variance between schools (23%). As a result of these comparisons, Model 9 was selected as the foundational level-1 model for further examination of the relationships between level-2 predictors and math achievement.

Table 38.
*Parameter Estimates for Combined Level-1 Predictors Model for Canada*

| Model | Type | Parameters | Estimates | *SE* | *p* | Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|
| 9 | Fixed | INT | 464.55 | 6.19 | <.001 | | | |
| | | Gender | 1.29 | 1.65 | .435 | | | |

Table 38.
*Parameter Estimates for Combined Level-1 Predictors Model for Canada*

| Model | Type | Parameters | Estimates | *SE* | *p* | Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|
| | | Extra lessons | -12.89 | 1.17 | <.001 | | | |
| | | Self-confidence | 35.79 | 1.20 | <.001 | | | |
| | | Valuing math | 4.51 | 1.28 | .001 | | | |
| | | Home resources | 2.90 | 1.76 | .100 | | | |
| | Random | $\tau_{00}$ | 1236.43 | 35.16 | <.001 | | | |
| | | Gender | 68.64 | 8.28 | .005 | | | |
| | | Extra lessons | 44.42 | 6.66 | <.001 | | | |
| | | Self-confidence | 58.44 | 7.64 | <.001 | | | |
| | | $\sigma^2$ | 1665.98 | 40.82 | | | | |
| | Pseudo $R^2$ | | | | | 7 | 0.11 | -0.01 |
| | | | | | | 8 | 0.23 | 0.37 |

Note: Pseudo $R^2$ refers to the difference in the proportion of variance between Model 9 and Models 7-8.

### *Research Question 3*

To what extent are instructional variables (i.e., opportunity to learn, activities in math lessons, amount of homework assignment, and instructional time) associated with TIMSS 2003 eighth-grade math scores in each country?

In addressing this research question, a similar strategy for model building used in Research Question 1 was applied here. That is, each of the level-2 instructional practice variables was first added to the foundational level-1 model (Model 9) to make Models 10-13. Then, as a group, those variables with significant fixed effects in Models 10-13 were included in the combined instructional practices model, Model 14. It is important to note that in these models, the predictor home resources was excluded because both of its fixed and random effects were not statistically significant in earlier models. Also, all possible cross interactions between level-1 and level-2 predictors were allowed in these models. The results of Models 10-14 are presented in Tables 39-41.

As can be seen in Table 39, Model 10 with opportunity to learn math topics as level-2 predictors of math achievement yielded five statistically significant cross-level interactions: (1) opportunity to learn data by gender ($\gamma$ = .14, *SE* = .07, *p* = .034), (2) opportunity to learn algebra by extra math lessons ($\gamma$ = .10, *SE* = .04, *p* = .017), (3) opportunity to learn geometry by extra math lessons ($\gamma$ = -.12, *SE* = .06, *p* = .040), (4) opportunity to learn data by student self-confidence in learning math ($\gamma$ = .11, *SE* = .04, *p* = .004), and opportunity to learn measurement and student self-confidence in learning math ($\gamma$ = .13, *SE* = .05, *p* < .013).

Table 39 also showed that when amount of homework assignment, content-related activities and instructional practice-related activities in math lessons, and average number of math instructional hours per year were added to Models 11-13, no statistically significant cross-level interaction effects were detected. Of the level-2 main effects, two were found statistically significant: instructional practices-related activities in math lessons in Model 12 ($\gamma$ = 29.07, *SE* = 12.28, *p* = .019) and average math instructional hours per year in Model 13 ($\gamma$ = -.24, *SE* = .07, *p* = .001). This means that with every unit increase in instructional practices-related activities in math lessons, students' math scores were expected to increase by 29.07 points after adjusting for level-1 variables but not for other level-2 variables in the model. Surprisingly, however, with every unit increase in average math instructional hours per year, student math scores were expected to decrease by .24 points, after adjusting for level-1 variables but not level-2 variables in the model.

Table 39.

*Parameter Estimates for Level-2 Instructional Practices Models for Canada*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 10 | Fixed | INT | 432.42 | 32.12 | <.001 |
| | | Opportunity_algebra | -0.09 | 0.11 | .421 |
| | | Opportunity_data | -0.46 | 0.11 | <.001 |
| | | Opportunity_geometry | 0.03 | 0.14 | .859 |
| | | Opportunity_measurement | 0.04 | 0.18 | .839 |

Table 39.
*Parameter Estimates for Level-2 Instructional Practices Models for Canada*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Opportunity_number | 0.75 | 0.31 | .017 |
| | | Gender | 8.34 | 17.75 | .638 |
| | | Opportunity_algebra*Gender | 0.02 | 0.06 | .769 |
| | | Opportunity_data*Gender | 0.14 | 0.07 | .034 |
| | | Opportunity_geometry*Gender | 0.02 | 0.08 | .817 |
| | | Opportunity_measurement*Gender | -0.08 | 0.09 | .352 |
| | | Opportunity_number*Gender | -0.13 | 0.18 | .464 |
| | | Extra lessons | -5.82 | 14.49 | .688 |
| | | Opportunity_algebra*Extra lessons | 0.10 | 0.04 | .017 |
| | | Opportunity_data*Extra lessons | 0.03 | 0.04 | .494 |
| | | Opportunity_geometry*Extra lessons | -0.12 | 0.06 | .040 |
| | | Opportunity_measurement*Extra lessons | 0.05 | 0.05 | .339 |
| | | Opportunity_number*Extra lessons | -0.11 | 0.14 | .456 |
| | | Self-confidence | 29.35 | 11.09 | .009 |
| | | Opportunity_algebra*Self-confidence | 0.03 | 0.04 | .421 |
| | | Opportunity_data*Self-confidence | 0.11 | 0.04 | .004 |
| | | Opportunity_geometry*Self-confidence | -0.06 | 0.04 | .168 |
| | | Opportunity_measurement*Self-confidence | 0.13 | 0.05 | .013 |
| | | Opportunity_number*Self-confidence | -0.09 | 0.10 | .368 |
| | | Valuing math | 4.63 | 1.33 | .001 |
| | Random | $\tau_{00}$ | 991.40 | 31.49 | <.001 |
| | | Gender | 71.93 | 8.48 | .009 |
| | | Extra lessons | 36.52 | 6.04 | .001 |
| | | Self-confidence | 28.24 | 5.31 | .048 |
| | | $\sigma^2$ | 1663.55 | 40.79 | |
| 11 | Fixed | INT | 481.13 | 6.15 | <.001 |
| | | Homework assignment | -7.53 | 5.09 | .140 |
| | | Gender | -4.02 | 3.98 | .314 |
| | | Homework assignment*Gender | 4.57 | 3.07 | .138 |
| | | Extra lessons | -17.22 | 2.67 | <.001 |
| | | Homework assignment*Extra lessons | 3.71 | 1.99 | .064 |
| | | Self-confidence | 33.74 | 2.15 | <.001 |
| | | Homework assignment*Self-confidence | 1.80 | 1.69 | .289 |
| | | Valuing math | 4.75 | 1.30 | <.001 |
| | Random | $\tau_{00}$ | 1237.67 | 35.18 | <.001 |
| | | Gender | 56.97 | 7.55 | .011 |
| | | Extra lessons | 39.96 | 6.32 | <.001 |
| | | Self-confidence | 58.18 | 7.63 | <.001 |
| | | $\sigma^2$ | 1667.94 | 40.84 | |
| 12 | Fixed | INT | 426.43 | 31.18 | <.001 |
| | | Content_activities | -3.22 | 17.24 | .852 |
| | | Instruction_activities | 29.07 | 12.28 | .019 |
| | | Gender | 10.59 | 14.70 | .472 |
| | | Content_activities*Gender | 3.05 | 8.38 | .716 |

Table 39.
*Parameter Estimates for Level-2 Instructional Practices Models for Canada*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Instruction_activities*Gender | -8.13 | 6.49 | .212 |
| | | Extra lessons | -13.95 | 12.67 | .272 |
| | | Content_activities*Extra lessons | -6.58 | 7.03 | .351 |
| | | Instruction_activities*Extra lessons | 6.90 | 5.50 | .211 |
| | | Self-confidence | 38.56 | 11.21 | .001 |
| | | Content_activities*Self-confidence | 4.96 | 5.57 | .375 |
| | | Instruction_activities*Self-confidence | -6.28 | 4.93 | .204 |
| | | Valuing math | 4.58 | 1.29 | .001 |
| | Random | $\tau_{00}$ | 1213.95 | 34.84 | <.001 |
| | | Gender | 66.64 | 8.16 | .006 |
| | | Extra lessons | 45.73 | 6.76 | <.001 |
| | | Self-confidence | 57.35 | 7.57 | <.001 |
| | | $\sigma^2$ | 1666.73 | 40.83 | |
| 13 | Fixed | INT | 473.11 | 3.37 | <.001 |
| | | Instructional hours | -0.24 | 0.07 | .001 |
| | | Gender | 1.12 | 1.66 | .500 |
| | | Instructional hours*Gender | 0.08 | 0.03 | .017 |
| | | Extra lessons | -12.89 | 1.15 | <.001 |
| | | Instructional hours*Extra lessons | 0.00 | 0.02 | .988 |
| | | Self-confidence | 35.71 | 1.21 | <.001 |
| | | Instructional hours*Self-confidence | 0.04 | 0.03 | .128 |
| | | Valuing math | 4.74 | 1.31 | .001 |
| | Random | $\tau_{00}$ | 1146.24 | 33.86 | <.001 |
| | | Gender | 54.42 | 7.38 | .013 |
| | | Extra lessons | 45.24 | 6.73 | <.001 |
| | | Self-confidence | 56.97 | 7.55 | <.001 |
| | | $\sigma^2$ | 1667.79 | 40.84 | |

In terms of model fit, in comparison with the foundational level-1 model (Model 9), Model 10 appeared to be the most efficient model because the amount of explained variance between schools in this model increased by 20% (see Table 40). As for the within school variance, no significant difference was observed between Models 10-13 and Model 9 (pseudo $R^2 = 0$).

Table 40.
*Comparison of $R^2$ between Level-2 Instructional Practice Models and Foundational Level-1 Model for Canada*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 10 vs. 9 | 0.20 | 0.00 |
| 11 vs. 9 | 0.00 | 0.00 |
| 12 vs. 9 | 0.02 | 0.00 |
| 13 vs. 9 | 0.07 | 0.00 |

Similar to Model 10, when using all the level-2 instructional practice variables to predict math achievement, Model 14 produced five statistically significant cross-level interaction effects (see Table 41). First, average math instructional hours per year interacted with gender ($\gamma$ = .08, *SE* = .04, *p* = .030). Second, opportunity to learn algebra interacted with extra math lessons ($\gamma$ = .11, *SE* = .05, *p* = .020). Third, opportunity to learn geometry interacted with extra math lessons ($\gamma$ = -.12, *SE* = .06, *p* = .029). Fourth, opportunity to learn data interacted with self-confidence in learning math ($\gamma$ = .11, *SE* = .04, *p* = .002). Finally, opportunity to learn measurement interacted with self-confidence in learning math ($\gamma$ = .14, *SE* = .05, *p* = .007). In this model, all the random effects were statistically significant, suggesting that, in Canada, the relationships between math achievement and gender, extra math lessons, and self-confidence in learning math differed significantly across schools.

Table 41.
*Parameter Estimates for the Combined Level-2 Instructional Practices Model for Canada*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 14 | Fixed | INT | 338.66 | 41.43 | <.001 |
| | | Homework assignment | -1.79 | 4.48 | .688 |
| | | Opportunity_algebra | -0.04 | 0.10 | .695 |
| | | Opportunity_data | -0.46 | 0.11 | <.001 |
| | | Opportunity_geometry | 0.05 | 0.13 | .687 |
| | | Opportunity_measurement | -0.01 | 0.17 | .975 |
| | | Opportunity_number | 0.95 | 0.32 | .004 |
| | | Content_activities | 30.23 | 16.22 | .063 |
| | | Instruction_activities | 13.98 | 10.83 | .198 |
| | | Instructional hours | -0.23 | 0.07 | .002 |

Table 41.
*Parameter Estimates for the Combined Level-2 Instructional Practices Model for Canada*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Gender | 30.92 | 24.51 | .208 |
| | | Homework assignment*Gender | 4.23 | 2.93 | .150 |
| | | Opportunity_algebra*Gender | 0.01 | 0.06 | .905 |
| | | Opportunity_data*Gender | 0.12 | 0.06 | .060 |
| | | Opportunity_geometry*Gender | 0.01 | 0.07 | .931 |
| | | Opportunity_measurement*Gender | -0.08 | 0.09 | .355 |
| | | Opportunity_number*Gender | -0.18 | 0.19 | .339 |
| | | Content_activities*Gender | -6.29 | 9.50 | .508 |
| | | Instruction_activities*Gender | -5.09 | 7.09 | .473 |
| | | Instructional hours*Gender | 0.08 | 0.04 | .030 |
| | | Extra lessons | 1.85 | 18.38 | .920 |
| | | Homework assignment*Extra lessons | 2.71 | 1.72 | .116 |
| | | Opportunity_algebra*Extra lessons | 0.11 | 0.05 | .020 |
| | | Opportunity_data*Extra lessons | 0.03 | 0.04 | .433 |
| | | Opportunity_geometry*Extra lessons | -0.12 | 0.06 | .029 |
| | | Opportunity_measurement*Extra lessons | 0.04 | 0.05 | .356 |
| | | Opportunity_number*Extra lessons | -0.18 | 0.15 | .229 |
| | | Content_activities*Extra lessons | -11.52 | 7.17 | .109 |
| | | Instruction_activities*Extra lessons | 8.51 | 5.24 | .105 |
| | | Instructional hours*Extra lessons | 0.00 | 0.02 | .909 |
| | | Self-confidence | 43.34 | 16.35 | .009 |
| | | Homework assignment*Self-confidence | -0.80 | 1.66 | .631 |
| | | Opportunity_algebra*Self-confidence | 0.03 | 0.04 | .425 |
| | | Opportunity_data*Self-confidence | 0.11 | 0.04 | .002 |
| | | Opportunity_geometry*Self-confidence | -0.06 | 0.04 | .137 |
| | | Opportunity_measurement*Self-confidence | 0.14 | 0.05 | .007 |
| | | Opportunity_number*Self-confidence | -0.11 | 0.11 | .315 |
| | | Content_activities*Self-confidence | -2.96 | 6.42 | .644 |
| | | Instruction_activities*Self-confidence | -3.73 | 4.38 | .396 |
| | | Instructional hours*Self-confidence | 0.03 | 0.02 | .223 |
| | | Valuing math | 4.43 | 1.33 | .001 |
| | Random | $\tau_{00}$ | 861.96 | 29.36 | <.001 |
| | | Gender | 60.11 | 7.75 | .025 |
| | | Extra lessons | 36.74 | 6.06 | .001 |
| | | Self-confidence | 28.88 | 5.37 | .042 |
| | | $\sigma^2$ | 1661.12 | 40.76 | |

As evident in Table 42, compared to previously constructed models (Models 9-13), the amounts of explained variance between schools in Model 14 were more significant. For example, an increase of 30% in the between school variance was observed when using Model 14 instead of Models 9 or 11. At minimum, changing from

Models 10 to Model 14 would result in 13% more of the variance between schools to be accounted for. However, for the variance within schools, no change was noted across these models. Thus, in consideration of the amount of the explained variance between schools, Model 14 surpassed previously constructed models in predicting math achievement.

Table 42.
*Comparison of $R^2$ between Model 14 and Previously Constructed Models 9-13 for Canada*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 14 vs. 9 | 0.30 | 0.00 |
| 14 vs. 10 | 0.13 | 0.00 |
| 14 vs. 11 | 0.30 | 0.00 |
| 14 vs. 12 | 0.29 | 0.00 |
| 14 vs. 13 | 0.25 | 0.00 |

The modeled means of predicted math achievement for the five statistically significant interactions resulted from Model 14 are displayed in Figures 28-32. The data in Figure 28 suggested that there was an inverse relationship between average math instructional hours per year and math achievement and that this relationship was different across female and male groups. Noticeably, regardless of average math instructional hours per year, female students appeared to outperform male students in math achievement. However, with low average math instructional hours per years, there was a small gap in math achievement between female and male students. As the average math instructional hours per year increased, the math achievement gap between female and male students became larger.

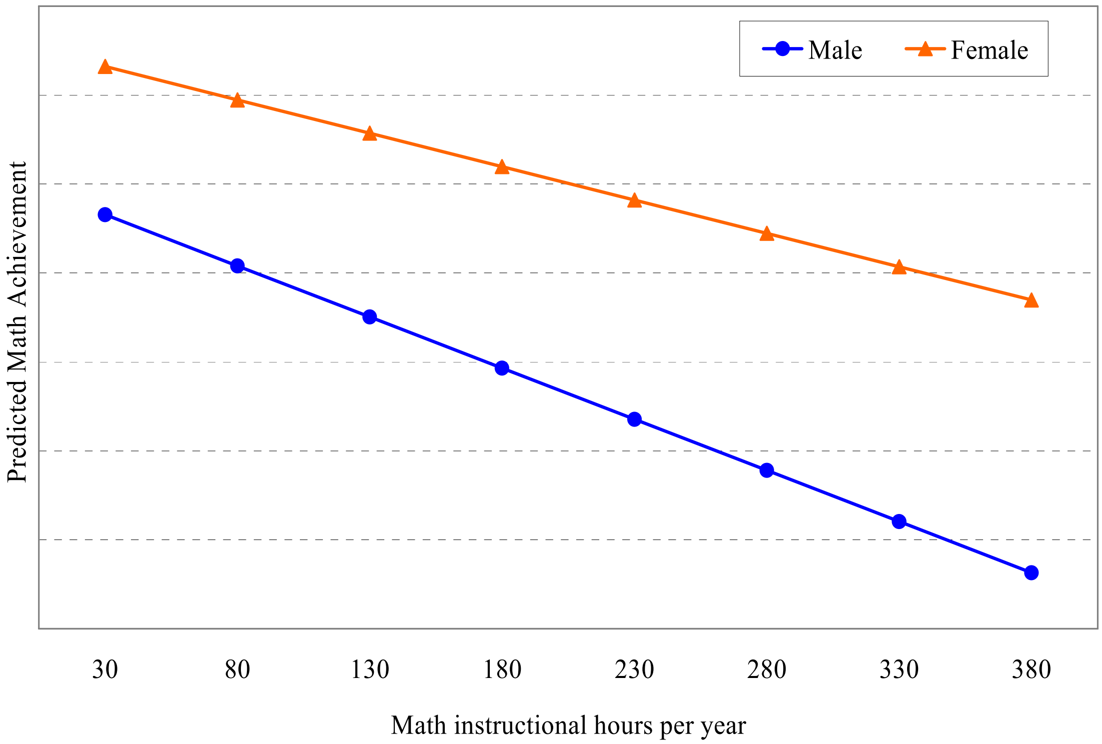



*Figure 28.* Interaction between Average Math Instructional Hours per Year and Gender for Canada

The data in Figure 29 depict the interaction between extra math lessons and opportunity to learn algebra. It appeared that when there was little opportunity to learn algebra, students tended to score similarly low in math, regardless of how frequently they took extra math lessons. However, the achievement gaps among students with different levels of extra math lessons grew rapidly as the opportunity to learn algebra increased. Specifically, students who reported taking extra math lessons everyday tended to achieve higher math scores than their peers who reported taking extra math lessons only sometimes or once or twice a week. As for the students who reported never taking extra math lessons, their math achievement seemed to decrease slightly when they had more opportunity to learn algebra.

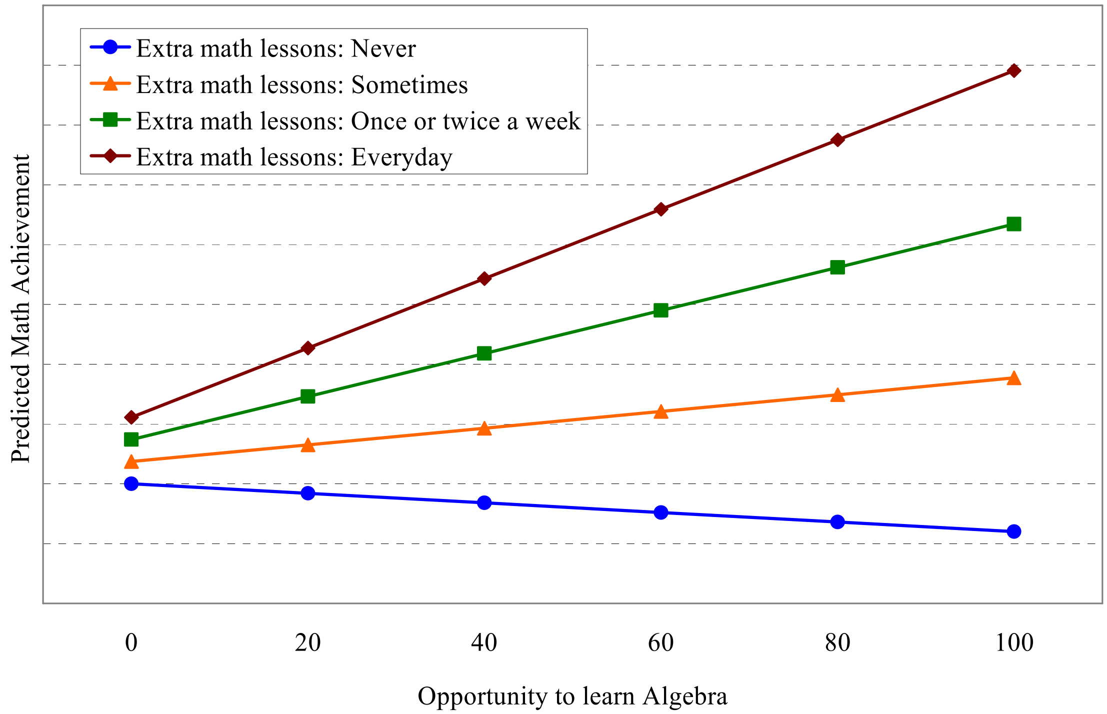



*Figure 29.* Interaction between Opportunity to Learn Algebra and Extra Math Lessons for Canada

The nature of the interaction between extra math lessons and opportunity to learn geometry is displayed in Figure 30. The results suggest that for students who reported never taking extra math lessons, increases in the opportunity to learn geometry was associated with increased math scores. However, as the frequencies of extra math lessons increased, the relationship rapidly became an inverse association and increased in the opportunity to learn geometry was associated with lower math scores. As an example, students who reported taking extra math lessons everyday tended to score lowest in math when the opportunity to learn geometry was the highest.

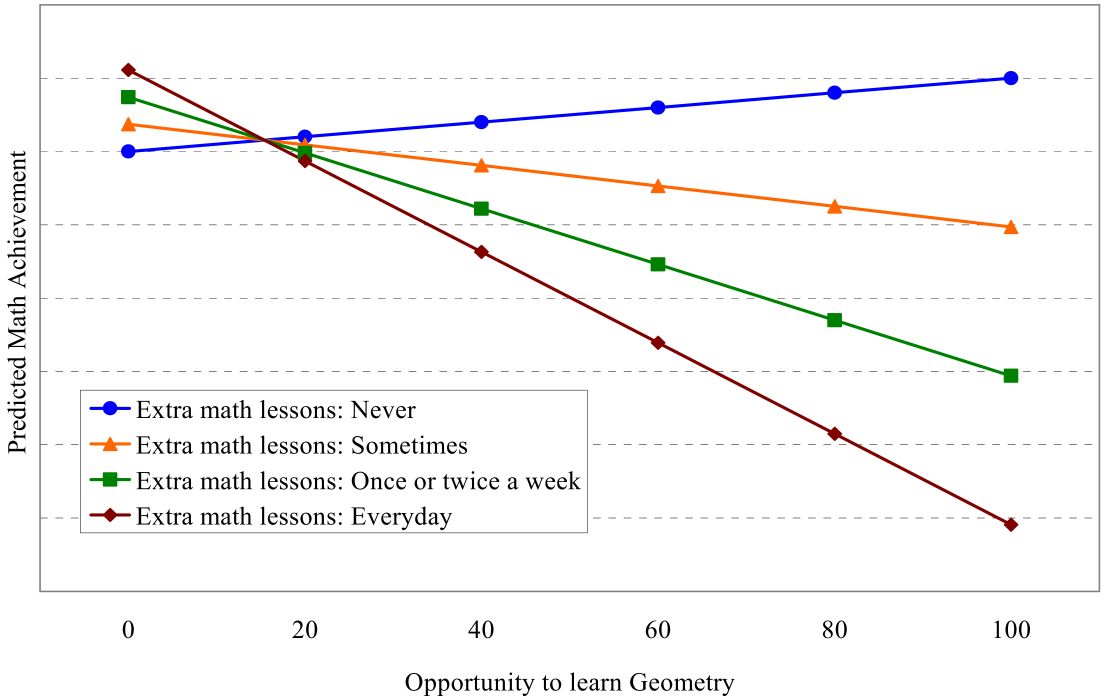



*Figure 30.* Interaction between Opportunity to Learn Geometry and Extra Math Lessons for Canada

Figure 31 presents the modeled mean of math achievement based upon student self-confidence in learning math and opportunity to learn data. Overall, it was noted that increases in opportunity to learn data were associated with decreases in student math scores, regardless of levels of student self-confidence in learning math. However, in comparing the three groups of students, those reported having a high level of self-confidence in learning math consistently outperformed their peers who reported having a low or medium level of self-confidence in learning math. The sizes of differences in math achievement among these three groups of students, however, was small when there was little opportunity to learn data and became slightly bigger when there was higher opportunity to learn data.

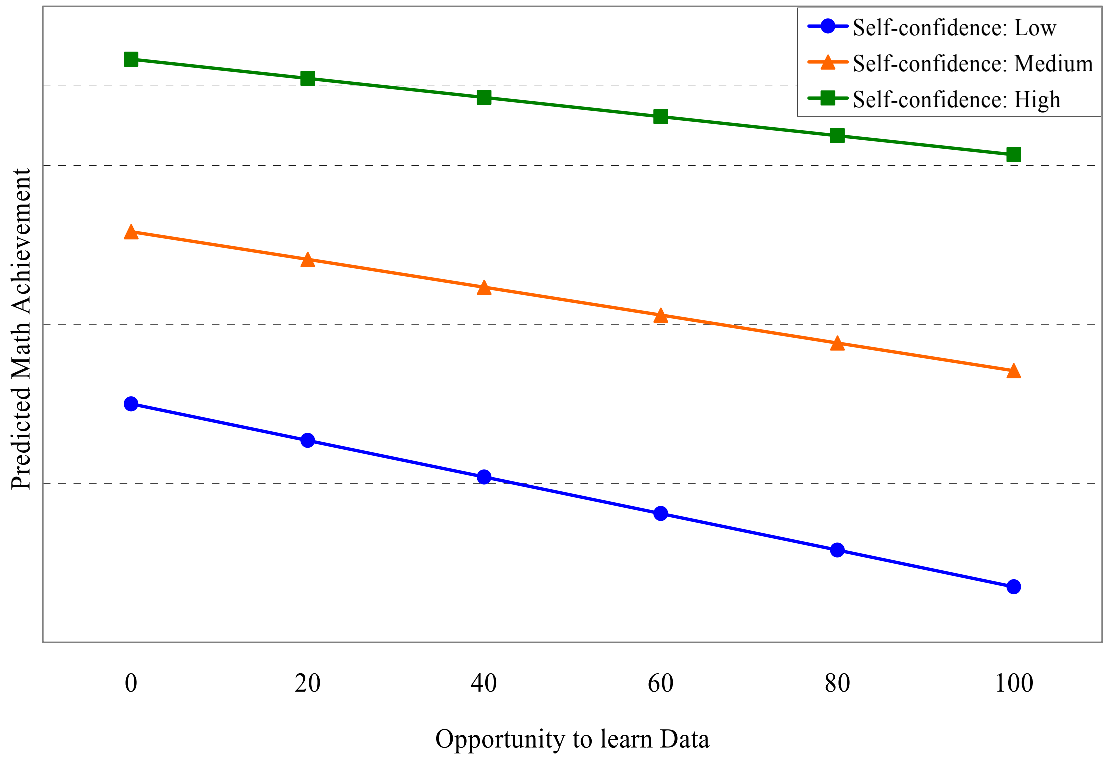



*Figure 31.* Interaction between Opportunity to Learn Data and Self-confidence for Canada

The interaction between opportunity to learn measurement and student self-confidence in learning math is illustrated in Figure 32. The results suggested that, for students with a medium or high level of self-confidence in learning math, increases in opportunity to learn measurement was associated with higher math scores; whereas for students with a low level of self-confidence in learning math, increases in opportunity to learn measurement made no difference in their math scores. Thus, the achievement gaps among these students was smallest when there was little opportunity to learn measurement and largest when there was high opportunity to learn measurement. As expected, regardless of opportunity to learn measurement, those students who had a higher level of self-confidence in learning math consistently performed better than their peers who had a lower level of self-confidence in learning math.

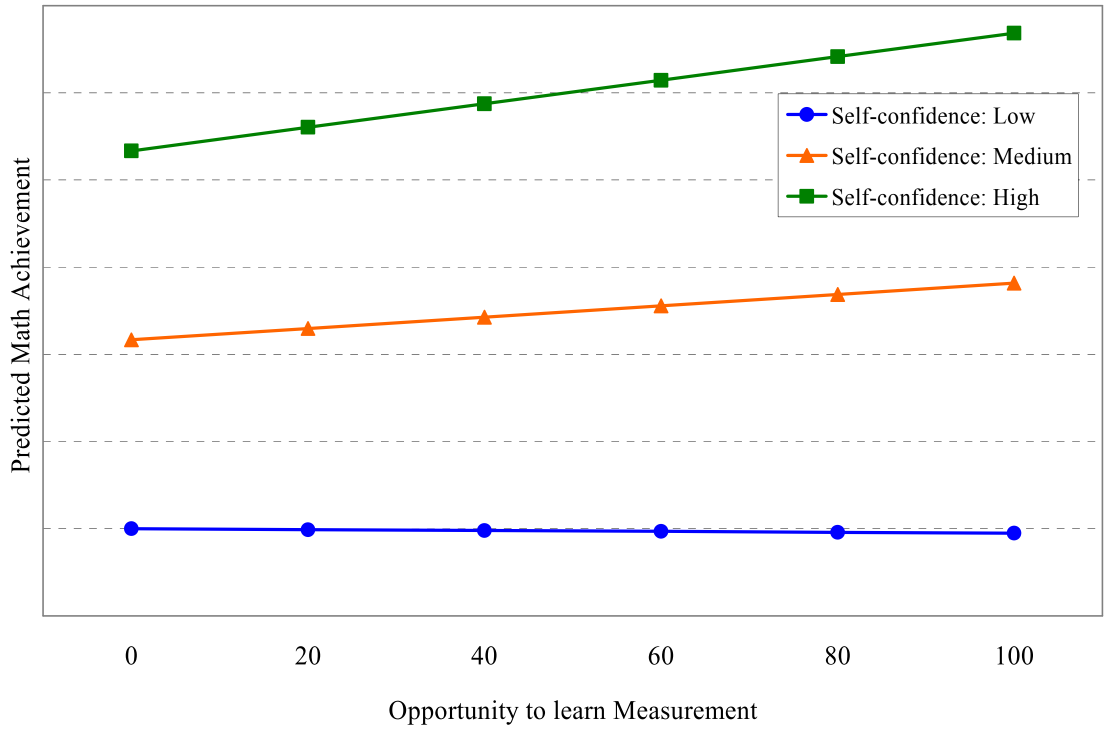



*Figure 32.* Interaction between Opportunity to Learn measurement and Self-confidence for Canada

*Research Question 4*

To what extent are teacher-related variables (i.e., preparation to teach, ready to teach, and professional development) associated with TIMSS 2003 eighth-grade math scores in each country?

Similarly, incremental model building strategies were applied to examine the relationships among teacher-related variables (i.e., preparation to teach, ready to teach, and professional development) and math achievement. Results of these models (Models 15-18) are presented in Tables 43-46.

Interestingly, as shown in Table 43, self-confidence in learning math was the only level-1 predictor that had statistically significant interaction with the level-2 predictors in Models 15-17. Specifically, in Model 15, student self-confidence in learning math interacted with math-related preparation to teach ($\gamma$ = -9.40, *SE* = 2.19, *p* <.001). Similarly, in Model 16 self-confidence in learning math interacted with ready to teach data ($\gamma$ = 6.52, *SE* = 3.18, *p* = .041). Likewise, in Model 17, self-confidence in learning math interacted with math-related professional development ($\gamma$ = 1.64, *SE* = 0.51, *p* = .002). Also, in these models, it was noted that all of the random effects were statistically significant, suggesting that a significant amount of variance in math achievement remained unexplained, both within and between schools.

Table 43.
*Parameter Estimates for Teacher Background Models for Canada*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 15 | Fixed | INT | 462.47 | 3.99 | <.001 |
| | | Preparation | 27.27 | 6.63 | <.001 |
| | | Gender | 2.26 | 2.29 | .325 |
| | | Preparation*Gender | -2.39 | 3.13 | .447 |
| | | Extra lessons | -12.76 | 1.66 | <.001 |
| | | Preparation*Extra lessons | -0.18 | 2.14 | .933 |
| | | Self-confidence | 39.24 | 1.40 | <.001 |
| | | Preparation*Self-confidence | -9.40 | 2.19 | <.001 |
| | | Valuing math | 4.79 | 1.30 | <.001 |
| | Random | $\tau_{00}$ | 1086.90 | 32.97 | <.001 |
| | | Gender | 70.54 | 8.40 | .006 |
| | | Extra lessons | 43.58 | 6.60 | <.001 |
| | | Self-confidence | 36.88 | 6.07 | .003 |
| | | $\sigma^2$ | 1667.27 | 40.83 | |
| 16 | Fixed | INT | 473.05 | 32.37 | <.001 |
| | | Ready_number | -13.23 | 18.58 | .477 |
| | | Ready_algebra | -3.05 | 9.32 | .744 |
| | | Ready_measurement | 8.11 | 8.42 | .337 |
| | | Ready_geometry | 21.99 | 11.49 | .056 |
| | | Ready_data | -13.85 | 8.30 | .096 |
| | | Gender | -5.52 | 9.47 | .560 |
| | | Ready_number*Gender | 7.04 | 6.06 | .247 |
| | | Ready_algebra*Gender | 1.49 | 3.67 | .685 |
| | | Ready_measurement*Gender | -3.73 | 3.92 | .343 |
| | | Ready_geometry*Gender | -2.96 | 5.84 | .613 |
| | | Ready_data*Gender | 1.55 | 3.57 | .664 |
| | | Extra lessons | 0.34 | 5.95 | .955 |
| | | Ready_number*Extra lessons | 0.27 | 4.27 | .951 |
| | | Ready_algebra*Extra lessons | -3.40 | 2.82 | .228 |
| | | Ready_measurement*Extra lessons | 1.21 | 3.90 | .756 |

Table 43.

*Parameter Estimates for Teacher Background Models for Canada*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Ready_geometry*Extra lessons | -6.65 | 4.11 | .107 |
| | | Ready_data*Extra lessons | 1.42 | 2.41 | .555 |
| | | Self-confidence | 32.74 | 8.89 | <.001 |
| | | Ready_number*Self-confidence | 2.77 | 5.43 | .610 |
| | | Ready_algebra*Self-confidence | 0.19 | 3.19 | .954 |
| | | Ready_measurement*Self-confidence | -3.36 | 2.98 | .261 |
| | | Ready_geometry*Self-confidence | -4.49 | 3.90 | .251 |
| | | Ready_data*Self-confidence | 6.52 | 3.18 | .041 |
| | | Home resources | 4.79 | 1.30 | <.001 |
| | Random | $\tau_{00}$ | 1199.81 | 34.64 | <.001 |
| | | Gender | 75.20 | 8.67 | .004 |
| | | Extra lessons | 41.71 | 6.46 | <.001 |
| | | Self-confidence | 54.66 | 7.39 | <.001 |
| | | $\sigma^2$ | 1665.68 | 40.81 | |
| 17 | Fixed | INT | 491.98 | 6.09 | <.001 |
| | | Professional development | -6.48 | 1.67 | <.001 |
| | | Gender | 0.07 | 3.56 | .983 |
| | | Professional development*Gender | 0.45 | 0.98 | .648 |
| | | Extra lessons | -13.03 | 2.03 | <.001 |
| | | Professional development*Extra lessons | 0.06 | 0.61 | .917 |
| | | Self-confidence | 30.93 | 1.70 | <.001 |
| | | Professional development*Self-confidence | 1.64 | 0.51 | .002 |
| | | Valuing math | 4.66 | 1.30 | .001 |
| | Random | $\tau_{00}$ | 1103.66 | 33.22 | <.001 |
| | | Gender | 84.86 | 9.21 | .004 |
| | | Extra lessons | 46.81 | 6.84 | <.001 |
| | | Self-confidence | 48.12 | 6.94 | .001 |
| | | $\sigma^2$ | 1663.44 | 40.79 | |

When comparing the proportion of variance accounted for by Models 15-17 with that of the foundational level-1 model (Model 9), it appears that Model 15 was the most efficient one (see Table 44). As an example, whereas the inclusion of preparation to teach math content in Model 15 resulted in a reduction of 12% in the between school variance to be explained; the addition of ready to teach math contents in Model 16 resulted in a reduction of only 3% in the between school variance to be explained. No improvement in the within school variance was noted by use of these models.

Table 44.
*Comparison of $R^2$ between Level-2 Teacher Background and Foundational Level-1 Model for Canada*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 15 vs. 9 | 0.12 | 0.00 |
| 16 vs. 9 | 0.03 | 0.00 |
| 17 vs. 9 | 0.11 | 0.00 |

When including all the teacher-related variables (i.e., preparation to teach, ready to teach math topics, and math-related professional development) in Model 18 to predict math achievement, two statistically significant cross-level interaction effects were produced (see Table 45). Specifically, preparation to teach math content was found to interact with student self-confidence in learning math ($\gamma$ = -7.83, *SE* = 2.28, *p* <.001) and math-related professional development was found to interact with student self-confidence in learning math ($\gamma$ = 1.10, *SE* = .53, *p* = .038). Also, in this model, all the random effects were statistically significant, meaning that a considerable amount of variance remained to be explained within and between schools.

Table 45.
*Parameter Estimates for the Combined Teacher Background Model for Canada*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 18 | Fixed | INT | 477.11 | 30.46 | <.001 |
| | | Preparation | 20.95 | 6.49 | .002 |
| | | Professional development | -5.22 | 1.74 | .003 |
| | | Ready_number | -7.71 | 17.98 | .668 |
| | | Ready_algebra | -3.11 | 9.21 | .736 |
| | | Ready_measurement | 8.89 | 8.22 | .281 |
| | | Ready_geometry | 13.32 | 11.95 | .266 |
| | | Ready_data | -9.29 | 8.60 | .281 |
| | | Gender | -5.24 | 9.50 | .581 |
| | | Preparation*Gender | -1.90 | 3.68 | .606 |
| | | Professional development*Gender | 0.26 | 1.14 | .822 |
| | | Ready_number*Gender | 6.73 | 6.22 | .281 |
| | | Ready_algebra*Gender | 1.61 | 3.84 | .674 |
| | | Ready_measurement*Gender | -3.80 | 4.10 | .356 |
| | | Ready_geometry*Gender | -2.55 | 5.97 | .669 |
| | | Ready_data*Gender | 1.22 | 3.71 | .742 |
| | | Extra lessons | -0.96 | 5.83 | .870 |
| | | Preparation*Extra lessons | 1.18 | 2.34 | .614 |
| | | Professional development*Extra lessons | 0.41 | 0.63 | .511 |

Table 45.
*Parameter Estimates for the Combined Teacher Background Model for Canada*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Ready_number*Extra lessons | 0.81 | 4.33 | .852 |
| | | Ready_algebra*Extra lessons | -3.91 | 2.83 | .168 |
| | | Ready_measurement*Extra lessons | 0.56 | 3.96 | .888 |
| | | Ready_geometry*Extra lessons | -6.80 | 4.25 | .110 |
| | | Ready_data*Extra lessons | 2.02 | 2.48 | .417 |
| | | Self-confidence | 32.59 | 8.04 | <.001 |
| | | Preparation*Self-confidence | -7.83 | 2.28 | .001 |
| | | Professional development*Self-confidence | 1.10 | 0.53 | .038 |
| | | Ready_number*Self-confidence | 0.64 | 5.13 | .901 |
| | | Ready_algebra*Self-confidence | 0.39 | 3.05 | .899 |
| | | Ready_measurement*Self-confidence | -3.15 | 3.01 | .298 |
| | | Ready_geometry*Self-confidence | -1.56 | 3.95 | .693 |
| | | Ready_data*Self-confidence | 5.24 | 2.99 | .080 |
| | | Valuing math | 4.75 | 1.30 | <.001 |
| | Random | $\tau_{00}$ | 978.58 | 31.28 | <.001 |
| | | Gender | 81.04 | 9.00 | .003 |
| | | Extra lessons | 43.56 | 6.60 | <.001 |
| | | Self-confidence | 29.34 | 5.42 | .020 |
| | | $\sigma^2$ | 1666.44 | 40.82 | |

As evident in Table 46, Model 18 appeared to be more efficient than Models 9 and 15-17 in terms of the amount of variance accounted for between schools. Specifically, an increase of 10% to 21% in the between school variance was likely to result when using Model 18 as opposed to Models 9, 15, 16, or 17.

Table 46.
*Comparison of $R^2$ between Model 18 and Previously Constructed Models 9 and 15-17*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 18 vs. 9 | 0.21 | 0.00 |
| 18 vs. 15 | 0.10 | 0.00 |
| 18 vs. 16 | 0.18 | 0.00 |
| 18 vs. 17 | 0.11 | 0.00 |

As shown in Figure 33, the interaction between preparation to teach math content and student self-confidence in learning math suggests that students' math achievement was positively related their self-confidence in learning math. That is, the more self-confident students expressed in learning math the better they performed in math. And, this relationship was true for all the students in Canada, regardless if their teachers reported being prepared or not to teach math content. As expected, in comparing two groups of students, the one with teachers who were prepared to teach consistently achieved higher math scores than the other group of students whose teachers were not prepared to teach. The size of differences in math achievement between the two groups was large when students expressed low self-confidence in learning math and became narrower as their self-confidence in learning math increased.

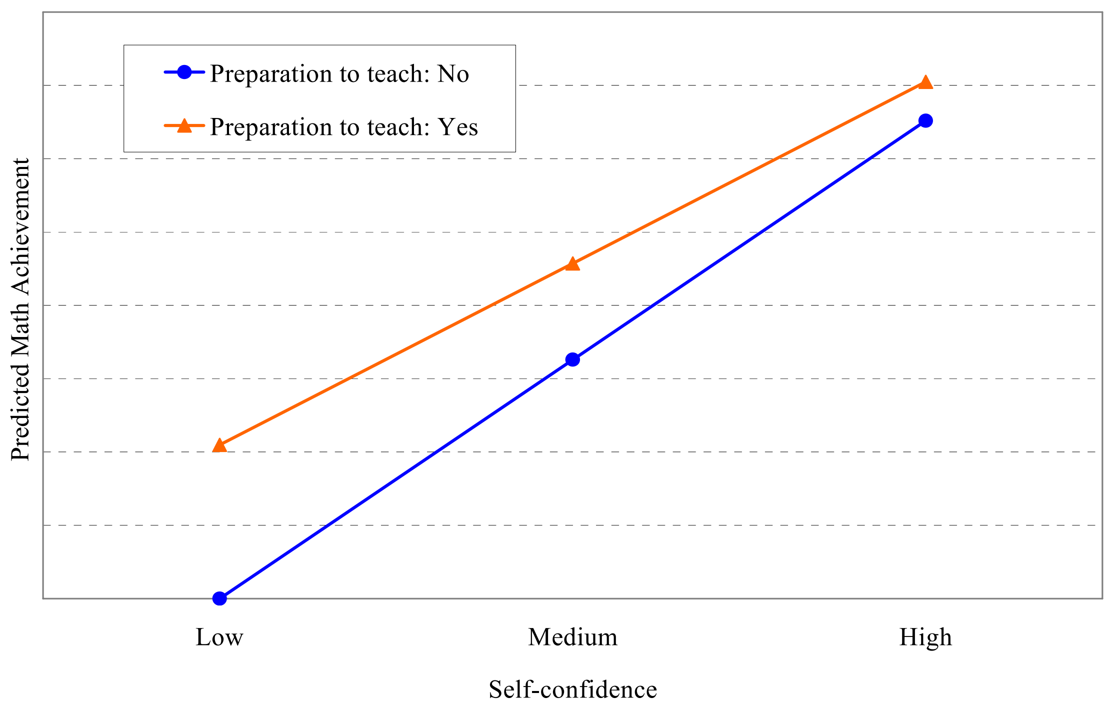



*Figure 33.* Interaction between Teacher Reported Preparation to Teach Math and Student Self-confidence in Learning Math for Canada

Figure 34 illustrates the interaction between student self-confidence in learning math and math-related professional development. Surprisingly, math achievement was found to be inversely associated with math related professional development. It seems that teachers' participation in more math-related professional development programs did not result in higher math performance for their students, regardless of how self-confident they were in learning math. It is important to note, however, that in comparing three groups of students, those with a higher self-confidence level in learning math consistently outperformed their peers who reported a lower level of self-confidence in learning math. The differences in their math achievement appeared to be largest when their teachers had five professional development programs and smallest when their teachers had none of these programs.

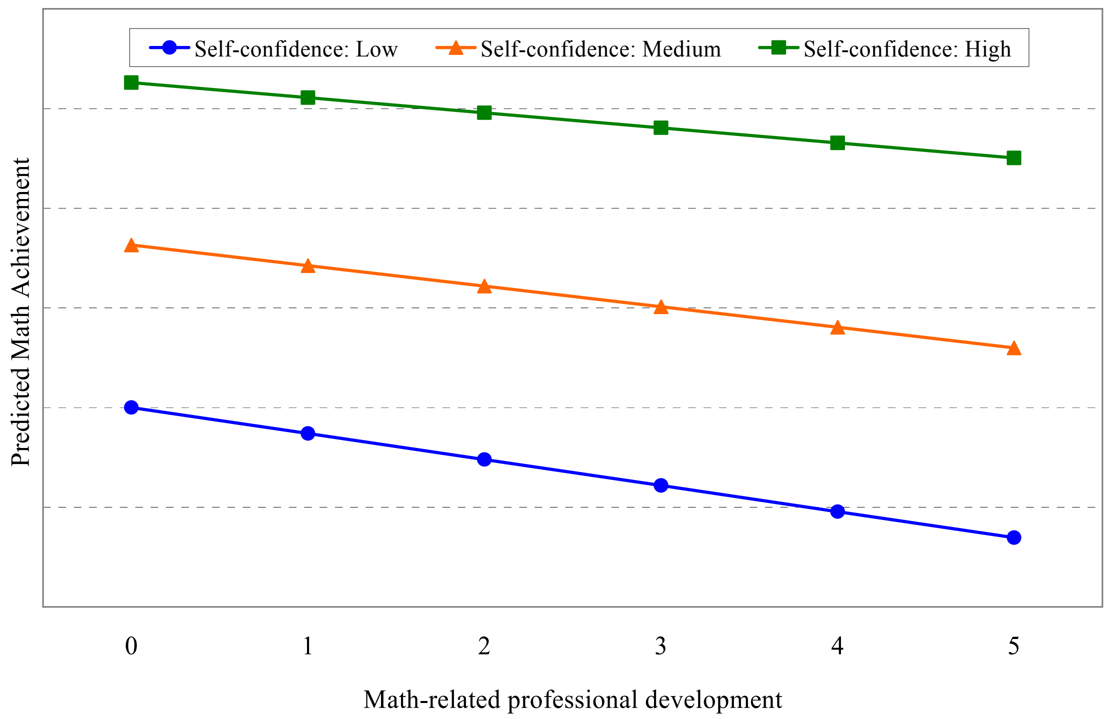



*Figure 34.* Interaction between Types of Math-Related Professional Development and Student Self-confidence in Learning Math for Canada

*Research Question 5*

To what extent are school-related variables (i.e., class size, school resources for math instruction, and teacher perception of math instructional limitation due to student factor) associated with TIMSS 2003 eighth-grade math scores in each country?

Table 47 provides a summary of the results for Models 19-21 where school-related variables (i.e., class size, school resources for math instruction, and teacher perception of math instructional limitations due to student factors) were separately included in the models to predict math achievement. In these models, no statistically significant cross-level interaction effects were detected. However, there were two statistically significant level-2 main effects: class size for math instruction in Model 19 ($\gamma$ = 15.69, *SE* = 6.24, *p* = .013) and teacher perception of math instructional limitations due to student factors in Model 20 ($\gamma$ = -17.84, *SE* = 4.83, *p* <.001). Also, in these models, all of the random effects were statistically significant, meaning that a good amount of variance remained to be explained within and between schools.

Table 47.
*Parameter Estimates for School Background Models for Canada*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 19 | Fixed | INT | 460.43 | 5.06 | <.001 |
| | | Class size | 15.69 | 6.24 | .013 |
| | | Gender | 3.21 | 3.27 | .328 |
| | | Class size*Gender | -2.34 | 3.19 | .463 |
| | | Extra lessons | -13.92 | 1.92 | <.001 |
| | | Class size*Extra lessons | 1.33 | 1.78 | .457 |
| | | Self-confidence | 37.72 | 2.12 | <.001 |
| | | Class size*Self-confidence | -2.40 | 2.18 | .272 |
| | | Valuing math | 4.78 | 1.30 | <.001 |
| | Random | $\tau_{00}$ | 1180.10 | 34.35 | <.001 |
| | | Gender | 64.75 | 8.05 | .006 |
| | | Extra lessons | 45.26 | 6.73 | <.001 |
| | | Self-confidence | 57.88 | 7.61 | <.001 |
| | | $\sigma^2$ | 1667.70 | 40.84 | |
| 20 | Fixed | INT | 481.84 | 4.12 | <.001 |

Table 47.
*Parameter Estimates for School Background Models for Canada*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Instructional limitation | -17.84 | 4.83 | <.001 |
| | | Gender | 0.33 | 1.91 | .864 |
| | | Instructional limitation*Gender | 1.92 | 2.51 | .444 |
| | | Extra lessons | -12.76 | 1.55 | <.001 |
| | | Instructional limitation*Extra lessons | -0.03 | 1.66 | .988 |
| | | Self-confidence | 35.36 | 1.35 | <.001 |
| | | Instructional limitation*Self-confidence | 0.73 | 1.67 | .663 |
| | | Valuing math | 4.83 | 1.30 | <.001 |
| | Random | $\tau_{00}$ | 1105.53 | 33.25 | <.001 |
| | | Gender | 65.91 | 8.12 | .006 |
| | | Extra lessons | 45.63 | 6.75 | <.001 |
| | | Self-confidence | 60.68 | 7.79 | <.001 |
| | | $\sigma^2$ | 1666.84 | 40.83 | |
| 21 | Fixed | INT | 459.82 | 7.73 | <.001 |
| | | School resources | 9.36 | 5.79 | .107 |
| | | Gender | -2.94 | 4.38 | .502 |
| | | School resources*Gender | 3.30 | 2.78 | .237 |
| | | Extra lessons | -16.01 | 3.80 | <.001 |
| | | School resources*Extra lessons | 2.37 | 2.42 | .329 |
| | | Self-confidence | 39.76 | 2.74 | <.001 |
| | | School resources*Self-confidence | -2.98 | 1.88 | .114 |
| | | Valuing math | 4.79 | 1.31 | <.001 |
| | Random | $\tau_{00}$ | 1243.31 | 35.26 | <.001 |
| | | Gender | 61.60 | 7.85 | .007 |
| | | Extra lessons | 43.69 | 6.61 | <.001 |
| | | Self-confidence | 52.70 | 7.26 | <.001 |
| | | $\sigma^2$ | 1669.78 | 40.86 | |

In comparing Models 19-21 with Model 9 in terms of the proportion of variance accounted for, it looks like that Model 20 worked the best (see Table 48). Specifically, the use of Model 20 as opposed to Model 9 increased the amount of between school variance accounted for by 11%; whereas the use of Model 19 as opposed to Model 9 accounted for only 5% more of the between school variance.

Table 48.
*Comparison of $R^2$ between Level-2 Teacher Background and Foundational Level-1 Model for Canada*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 19 vs. 9 | 0.05 | 0.00 |
| 20 vs. 9 | 0.11 | 0.00 |
| 21 vs. 9 | -0.01 | 0.00 |

Unlike earlier combined models, Model 22 with all of the school background-related predictors did not produce any statistically significant cross-level interaction effects. There were, however, two statistically significant level-2 main effects: class size for math instruction ($\gamma$ = 13.90, *SE* = 5.64, *p* = .015) and teacher perception of math instructional limitations due to student factors ($\gamma$ = -17.02, *SE* = 5.18, *p* = .002). These results suggest that, after controlling for all other level-1 and level-2 variables in the model, students in schools with larger class sizes tended to perform better in math than their peers in schools with smaller class sizes. Also, it seems that students tended to fare poorer in math in schools where teachers perceived to have more limitations due to student factors than in schools where teachers perceived to have none or few limitations due to student factors. Specifically, for every unit increased in class size, the students were expected to improve their math scores by 15.69 points, and for every unit increased in teacher perception of instructional limitations due to student factors, the students were expected to lower their math scores by 17.02 points, after controlling for other level-1 and level-2 variables in the model.

Again, in Model 22, all the random effects were statistically significant. This indicated that the amount of within and between school variance that remained to be explained were still significant.

Table 49.
*Parameter Estimates for the Combined School Background Model for Canada*

| Model | Type | Parameters | Estimates | *SE* | *P* |
|---|---|---|---|---|---|
| 22 | Fixed | INT | 460.40 | 9.16 | <.001 |
| | | Instructional limitation | -17.02 | 5.18 | .002 |
| | | Class size | 13.90 | 5.64 | .015 |
| | | School resources | 7.75 | 5.24 | .141 |
| | | Gender | -2.17 | 5.66 | .701 |
| | | Instructional limitation*Gender | 1.78 | 2.55 | .485 |
| | | Class size*Gender | -2.81 | 3.14 | .372 |

Table 49.
*Parameter Estimates for the Combined School Background Model for Canada*

| Model | Type | Parameters | Estimates | *SE* | *P* |
|---|---|---|---|---|---|
| | | School resources*Gender | 3.67 | 2.74 | .182 |
| | | Extra lessons | -16.55 | 4.11 | <.001 |
| | | Instructional limitation*Extra lessons | 0.14 | 1.63 | .931 |
| | | Class size*Extra lessons | 0.91 | 1.82 | .619 |
| | | School resources*Extra lessons | 2.25 | 2.46 | .362 |
| | | Self-confidence | 40.76 | 3.67 | <.001 |
| | | Instructional limitation*Self-confidence | 0.61 | 1.79 | .735 |
| | | Class size*Self-confidence | -2.12 | 2.21 | .339 |
| | | School resources*Self-confidence | -2.77 | 1.88 | .141 |
| | | Valuing math | 4.88 | 1.29 | <.001 |
| | Random | $\tau_{00}$ | 1042.32 | 32.28 | <.001 |
| | | Gender | 62.62 | 7.91 | .007 |
| | | Extra lessons | 45.34 | 6.73 | <.001 |
| | | Self-confidence | 57.70 | 7.60 | <.001 |
| | | $\sigma^2$ | 1667.59 | 40.84 | |

As shown in Table 50, Model 22 appears to be more efficient than Models 9 and 19-21. Although the amount of explained variance within schools in Model 22 did not change compared to these models (pseudo $R^2 = 0$), the amount of explained variance between schools in Model 22 increased by 6% (compared to Model 20) to 16% (compared to Models 9 and 21).

Table 50.
*Comparison of $R^2$ between Model 22 and Previously Constructed Models 9 and 19-21 for Canada*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 22 vs. 9 | 0.16 | 0.00 |
| 22 vs. 19 | 0.12 | 0.00 |
| 22 vs. 20 | 0.06 | 0.00 |
| 22 vs. 21 | 0.16 | 0.00 |

### *Final Model*

With an intention to identify the most efficient and parsimonious model to predict eighth-grade math achievement in Canada, Model 23 was built by including all the statistically significant level-2 predictors in earlier combined models to Model 9 and then compared with the three combined models, Models 14, 18, and 22.

As can be seen from Table 51, Model 23 produced six statistically significant cross-level interaction effects: (1) average math instructional hours per year by gender ( $\gamma$ = .09, SE = .04, p = .027), (2) opportunity to learn data by gender ( $\gamma$ = .14, *SE* = .07, *p* = .042), (3) opportunity to learn algebra by extra math lessons ( $\gamma$ = .10, *SE* = .04, *p* = .021), (4) opportunity to learn geometry by extra math lessons ( $\gamma$ = -.14, *SE* = .06, *p* = .018), (5) opportunity to learn data by self-confidence in learning math ( $\gamma$ = .08, *SE* = .04, *p* = .022). Finally, preparation to teach math content by self-confidence in learning math ( $\gamma$ = -5.85, *SE* = 2.20, *p* = .009). In addition, there were four significant level-2 main effects: (1) teacher perception of math instructional limitations due to student factors ( $\gamma$ = -14.68, *SE* = 4.02, *p* <.001), (2) opportunity to learn data ( $\gamma$ = -.26, *SE* = .11, *p* = .018), (3) math-related professional development ( $\gamma$ = -3.52, *SE* = 1.66, *p* = .035), and (4) average math instructional hours per year ( $\gamma$ = -.21, *SE* = .07, *p* = .005). The only level-1 variable that had significant main effect was student self-confidence in learning math ( $\gamma$ = 31.02, *SE* = 11.46, *p* = .008). Also, in this model, all of the random effects were statistically significant.

Table 51.
*Parameter Estimates for Full Model for Canada*

| Model | Type | Parameters | Estimates | *SE* | *P* |
|---|---|---|---|---|---|
| 23 | Fixed | INT | 446.85 | 28.36 | <.001 |
| | | Instructional limitation | -14.68 | 4.02 | .001 |
| | | Class size | 4.27 | 4.98 | .392 |
| | | Opportunity_Algebra | -0.11 | 0.10 | .246 |
| | | Opportunity_Data | -0.26 | 0.11 | .018 |
| | | Opportunity_Geometry | 0.00 | 0.14 | .986 |
| | | Opportunity_Measurement | 0.11 | 0.17 | .511 |
| | | Opportunity_number | 0.53 | 0.25 | .037 |
| | | Preparation | 16.89 | 6.08 | .006 |
| | | Professional development | -3.52 | 1.66 | .035 |
| | | Instructional hours | -0.21 | 0.07 | .005 |
| | | Gender | 16.17 | 19.42 | .406 |

Table 51.
*Parameter Estimates for Full Model for Canada*

| Model | Type | Parameters | Estimates | *SE* | *P* |
|---|---|---|---|---|---|
| | | Instructional limitation*Gender | 0.84 | 2.31 | .716 |
| | | Class size*Gender | 0.13 | 3.55 | .971 |
| | | Opportunity_algebra*Gender | -0.01 | 0.06 | .848 |
| | | Opportunity_data*Gender | 0.14 | 0.07 | .042 |
| | | Opportunity_geometry*Gender | -0.01 | 0.08 | .949 |
| | | Opportunity_measurement*Gender | -0.05 | 0.09 | .629 |
| | | Opportunity_number*Gender | -0.20 | 0.19 | .290 |
| | | Preparation*Gender | 0.68 | 3.60 | .851 |
| | | Professional development*Gender | -0.48 | 0.93 | .602 |
| | | Instructional hours*Gender | 0.09 | 0.04 | .027 |
| | | Extra lessons | -3.20 | 15.40 | .836 |
| | | Instructional limitation*Extra lessons | -0.04 | 1.57 | .978 |
| | | Class size*Extra lessons | 1.70 | 1.91 | .376 |
| | | Opportunity_algebra*Extra lessons | 0.10 | 0.04 | .021 |
| | | Opportunity_data*Extra lessons | 0.05 | 0.05 | .295 |
| | | Opportunity_geometry*Extra lessons | -0.14 | 0.06 | .018 |
| | | Opportunity_measurement*Extra lessons | 0.07 | 0.05 | .179 |
| | | Opportunity_number*Extra lessons | -0.15 | 0.15 | .301 |
| | | Preparation*Extra lessons | 1.39 | 2.33 | .552 |
| | | Professional development*Extra lessons | -0.41 | 0.66 | .537 |
| | | Instructional hours*Extra lessons | 0.00 | 0.02 | .993 |
| | | Self-confidence | 31.02 | 11.46 | .008 |
| | | Instructional limitation*Self-confidence | 0.27 | 1.48 | .856 |
| | | Class size*Self-confidence | 0.64 | 2.15 | .766 |
| | | Opportunity_algebra*Self-confidence | 0.04 | 0.04 | .302 |
| | | Opportunity_Data*Self-confidence | 0.08 | 0.04 | .022 |
| | | Opportunity_Geometry*Self-confidence | -0.06 | 0.04 | .168 |
| | | Opportunity_Measurement*Self-confidence | 0.10 | 0.05 | .065 |
| | | Opportunity_number*Self-confidence | -0.07 | 0.10 | .502 |
| | | Preparation*Self-confidence | -5.85 | 2.20 | .009 |
| | | Professional development*Self-confidence | 0.31 | 0.60 | .607 |
| | | Instructional hours*Self-confidence | 0.03 | 0.03 | .291 |
| | | Valuing math | 4.76 | 1.31 | .001 |
| | Random | $\tau_{00}$ | 710.83 | 26.66 | <.001 |
| | | Gender | 76.07 | 8.72 | .013 |
| | | Extra lessons | 40.05 | 6.33 | <.001 |
| | | Self-confidence | 24.51 | 4.95 | .072 |
| | | $\sigma^2$ | 1661.06 | 40.76 | |

As evident in Table 52, Model 23 appears to be the most efficient model for Canada because this model accounted for the largest amount of variance within and between schools. Specifically, Model 23 accounted for 18% more of the between school variance when compared to Model 14, and 27% when compared to Model 18 and 32% when compared to Model 22. Therefore, for Canada, Model 23 serves as the best model for predicting math achievement.

Table 52.
*Comparison of $R^2$ between Model 23 and Previously Constructed Models 14, 18 and 22 for Canada*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 14 | 0.18 | 0.00 |
| 18 | 0.27 | 0.00 |
| 22 | 0.32 | 0.00 |

Figure 35 visually displays the nature of the cross-level interactions between average math instructional hours per year and gender. The results suggested that there was an inverse relationship between average math instructional hours per year and math achievement and that this relationship was different across female and male groups. Noticeably, regardless of average math instructional hours per year, female students appeared to outperform male students in math achievement. However, with low average math instructional hours per years, there was a small gap in math achievement between female and male students. As the average math instructional hours per year increased, the math achievement gap between female and male students became more noticeable.

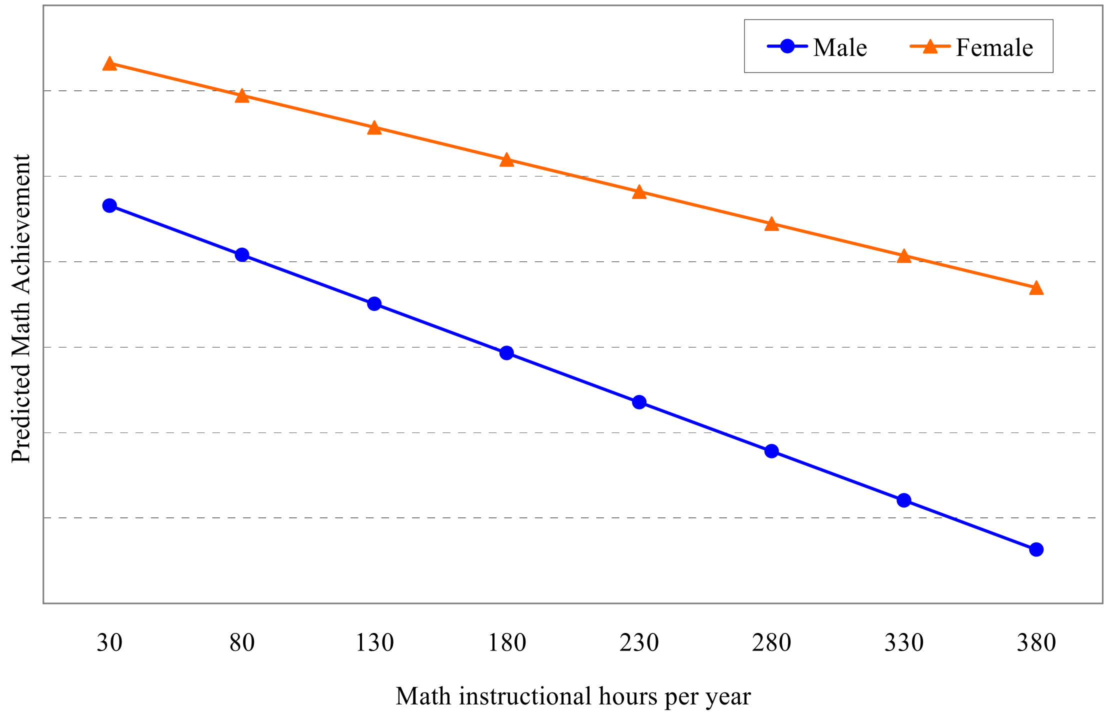



*Figure 35.* Interaction between Average Math Instructional Hours per Year and Gender for Canada

As shown in Figure 36, the interaction between preparation to teach math content and student self-confidence in learning math suggests that students' math achievement was positively related to their self-confidence in learning math. That is, the more self-confidence students expressed in learning math the better they performed in math. And, this relationship was true for all the students in Canada, regardless if their teachers reported being prepared or not to teach math content. As expected, in comparing two groups of students, the one with teachers who were prepared to teach consistently achieved higher math scores than the other group of students whose teachers were not prepared to teach. The size of differences in math achievement between the two groups was large when students expressed low self-confidence in learning math and became narrower as their self-confidence in learning math increased.

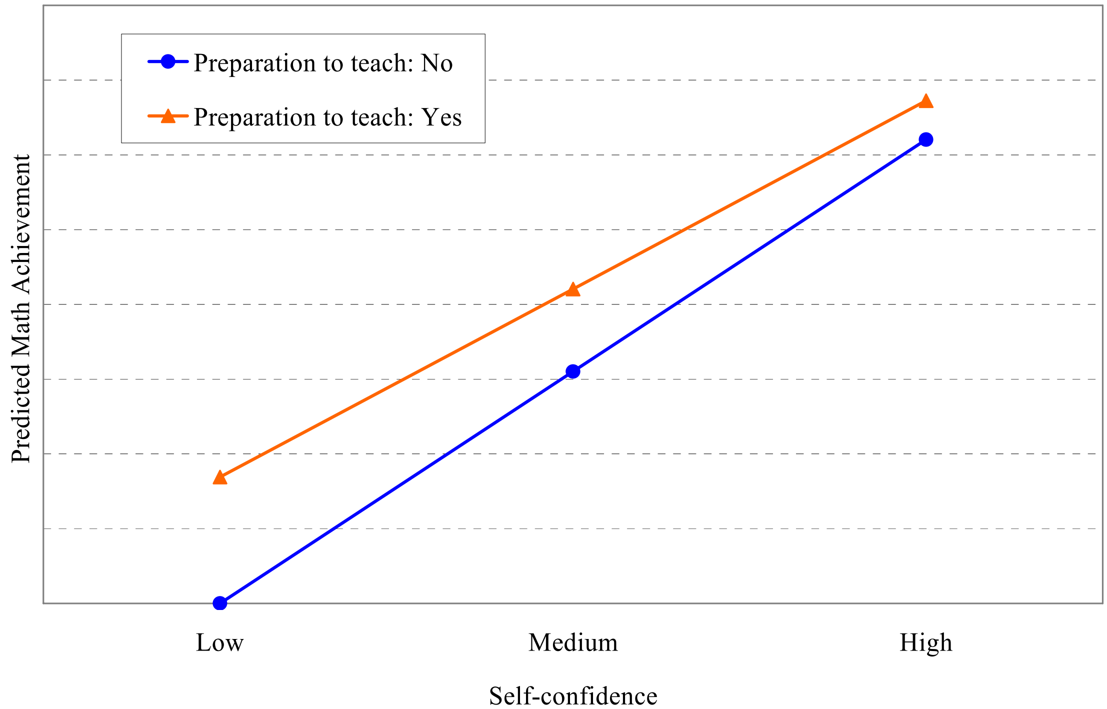



*Figure 36.* Interaction between Teacher Reported Preparation to Teach Math Content and Student Self-confidence in Learning Math for Canada

The nature of the cross-level interactions between opportunity to learn data and gender is presented in Figure 37. The data suggested that student math achievement was inversely related to opportunity to learn data and that this relationship differed significantly across female and male groups. Notably, regardless of opportunity to learn data, female students appeared to perform better in math than male students. However, with little opportunity to learn data, the achievement gap between female and male students seemed narrow. As the opportunity to learn data increased, the math achievement gaps between female and male students became statistically significant.

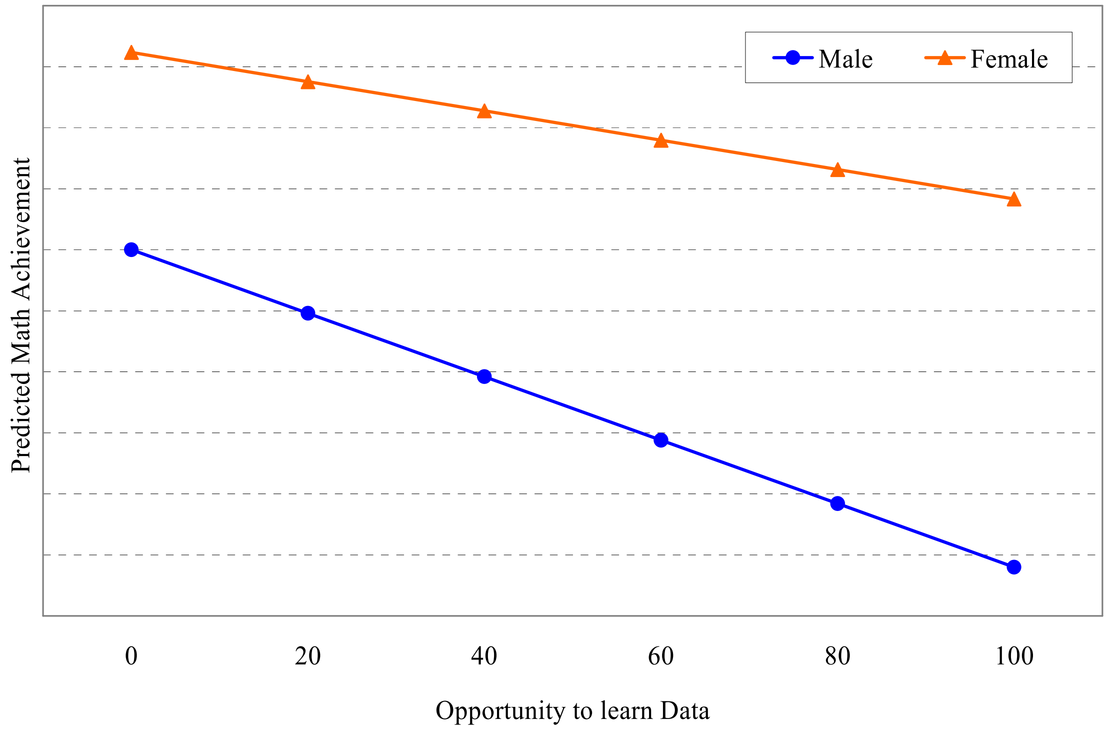



*Figure 37.* Interaction between Opportunity To Learn Data and Gender for Canada

Figure 38 presents the modeled mean of math achievement based upon student self-confidence in learning math and opportunity to learn data. Overall, it was noted that increases in opportunity to learn data were associated with decreases in student math scores, regardless of levels of student self-confidence in learning math. However, in comparing the three groups of students, those who reported having a high level of self-confidence in learning math consistently outperformed their peers who reported having a low or medium level of self-confidence in learning math. The sizes of differences in math achievement among these three groups of students, however, was small when there was little opportunity to learn data, and became slightly bigger as the opportunity to learn data increased.

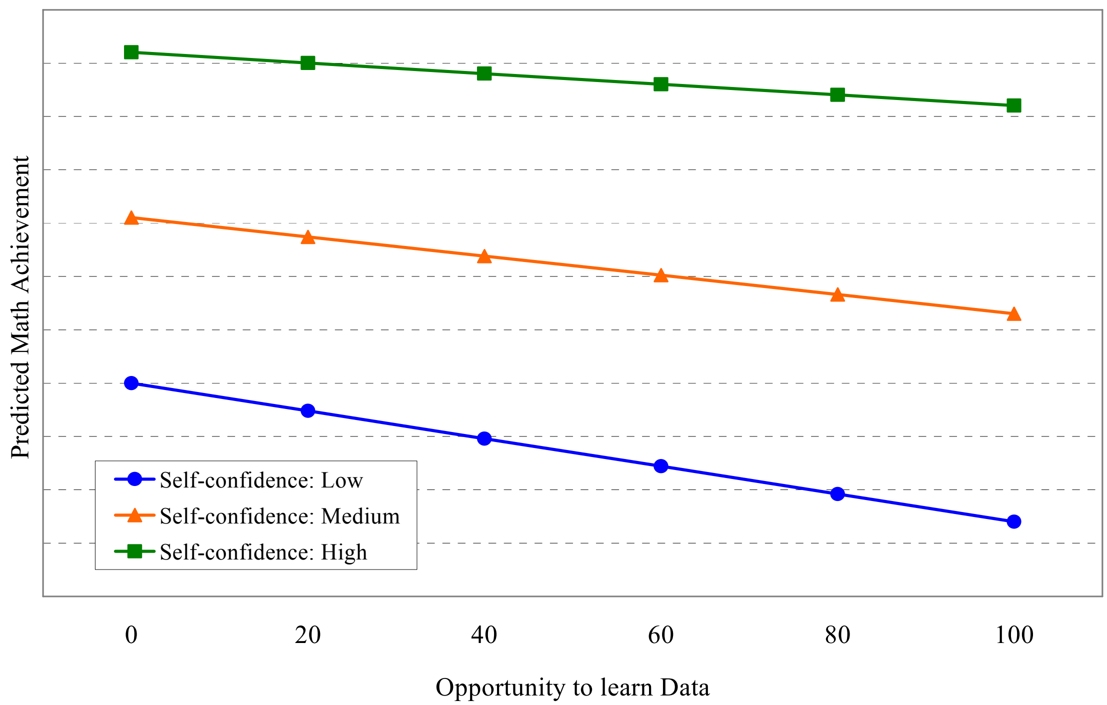



*Figure 38.* Interaction between Opportunity to Learn Data and Self-confidence in Learning Math for Canada

The nature of the interaction between extra math lessons and opportunity to learn geometry is displayed in Figure 39. The results suggest that for students who reported never taking extra math lessons, increases in the opportunity to learn geometry was associated with slightly increased math scores. However, as the frequencies of extra math lessons increased, the relationship rapidly became an inverse association and an increase in the opportunity to learn geometry was associated with lower math scores. As an example, students who reported taking extra math lessons everyday tended to score lowest in math when the opportunity to learn geometry was the highest. Also it was observed that when the opportunity to learn geometry was little the achievement gap across student groups was small. As the opportunity to learn geometry increased, the achievement gaps became significantly larger.

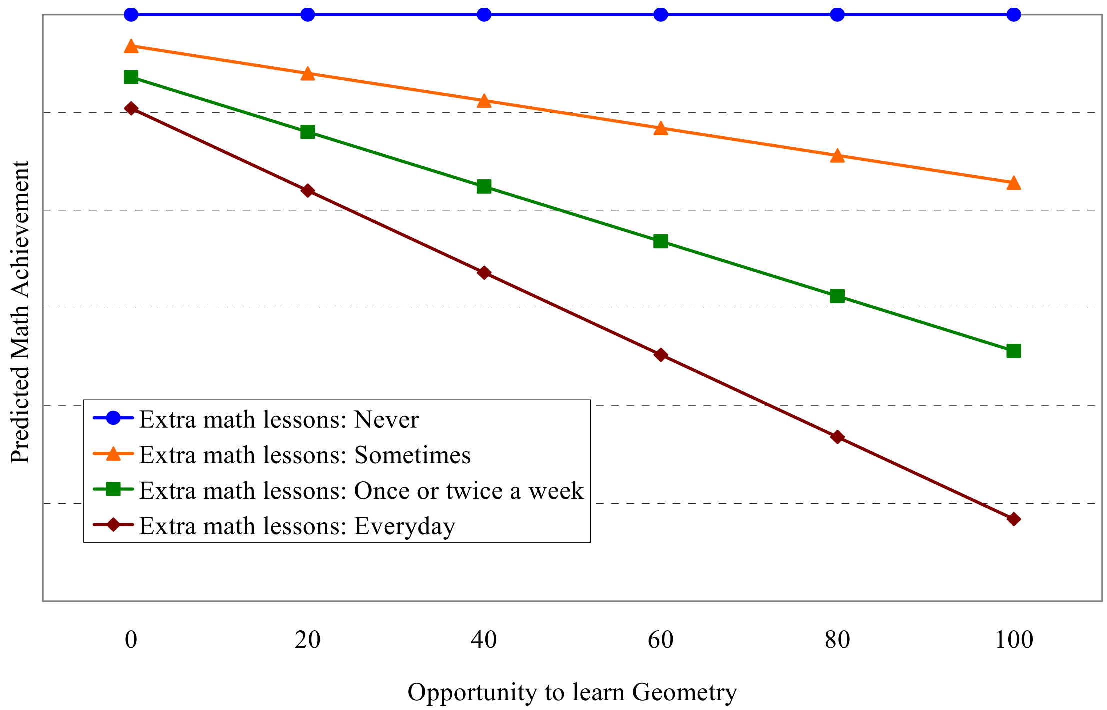



*Figure 39.* Interaction between Opportunity to Learn Geometry and Extra Math Lessons for Canada

Finally, the data in Figure 40 depict the interaction between extra math lessons and opportunity to learn algebra. It appeared that when there was little opportunity to learn algebra, students tended to score similarly low in math, regardless of how frequently they took extra math lessons. However, the achievement gaps among students with different levels of extra math lessons grew rapidly as the opportunity to learn algebra increased. Specifically, students who reported taking extra math lessons everyday tended to achieve higher math scores than their peers who reported taking extra math lessons only sometimes or once or twice a week. As for the students who reported never taking extra math lessons, their math achievement seemed to decrease slightly when they had more opportunity to learn algebra.

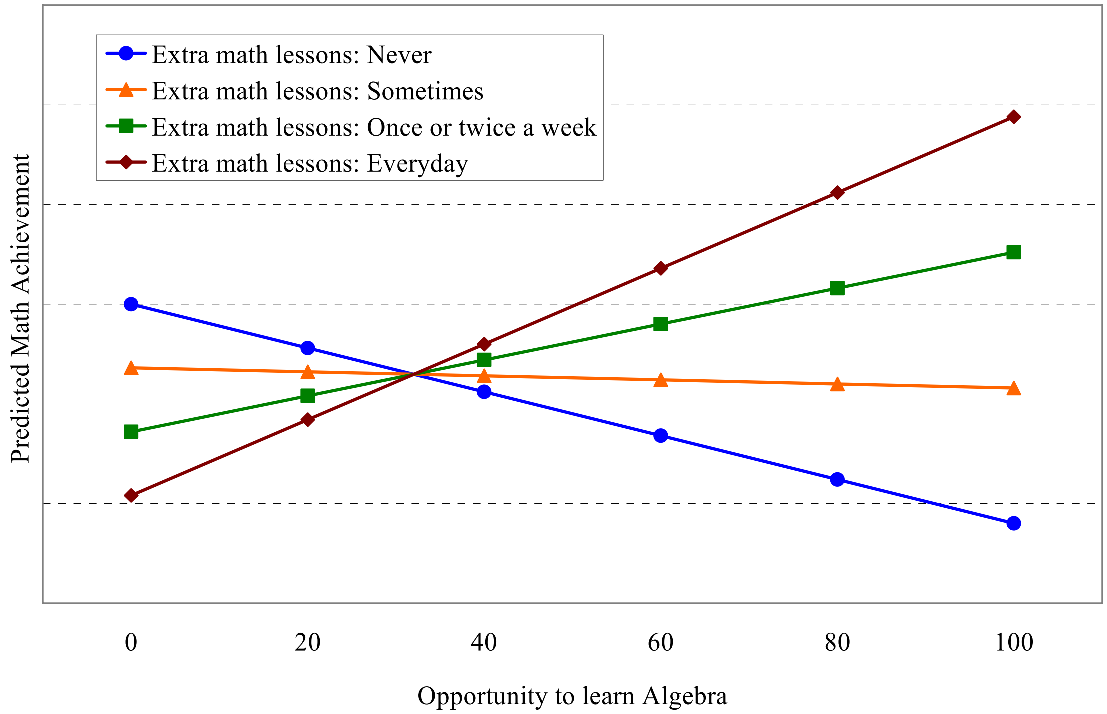



*Figure 40.* Interaction between Opportunity to Learn Algebra and Extra Math Lessons for Canada

## Results for Egypt

### *Evaluation of Missing Data*

As a result of the listwise deletion method, the sample size for Egypt was reduced from 7,095 students and 217 schools to 1,876 students and 69 schools. This means only 26.44% of the original sample had complete data on all variables of interest in this study. In order to evaluate the extent to which the data for Egypt were missing completely at random, the missingness on 19 level-2 variables were correlated. Results of this analysis suggested a non-randomness of missing data, with correlation coefficients ranging from .38 to .97 ($n$ = 217, $p$ <.001), indicating a modest to strong positive relationship among missingness indicators of the variables. In addition, when missingness was correlated with values of itself as well as values of other variables, only marginal correlations were observed ($r$ = -.26 to .14, $n$ = 217, $p$ <.001). In summary, the missing data mechanism for Egypt was not missing completely at random.

### *Univariate Analysis*

A descriptive examination of level-1 variables (i.e., overall math achievement, gender, self-confidence in learning math, valuing of math, time on math homework, extra math lessons, and home resources for learning math) was conducted using SAS 9.13. Of the complete sample of 1,876 eighth-grade students, 1,083 (57.73%) were male and 793 (42.27%) were female. On average, the weighted overall math achievement for Egyptian students was 416.52 (*SD* = 84.83) with the lowest score of 188.57 and the highest score of 714.09 (see Table 53).

With regard to level-1 predictor variables, it appeared that, on average, eighth-grade students in Egypt had a moderate support at home in terms of resources for learning (*M* = 1.94 , *SD* = .61), were above medium level of self-confidence in learning math (*M* = 1.50, *SD* = .62) and valuing of math (*M* = 1.79, *SD* = .61), spent a modest amount of time on math homework (*M* = .91, *SD* = .62), and took extra math lessons about one to two times a week (*M* = 1.70, *SD* = 1.02) (see Table 53).

Table 53.

*Weighted Descriptive Statistics for Level-1 Variables for Egypt (N = 1,876)*

| Variable | *M* | *SD* | Min | Max |
|---|---|---|---|---|
| Overall math achievement | 416.52 | 84.83 | 188.57 | 714.09 |
| Self-confidence in learning math | 1.50 | 0.62 | 0 | 2 |
| Valuing of math | 1.79 | 0.46 | 0 | 2 |
| Time on math homework | 0.91 | 0.62 | 0 | 2 |
| Extra math lessons | 1.70 | 1.02 | 0 | 3 |
| Home resources for learning math | 1.94 | 0.61 | 0 | 3 |

Note: When weight was used to compute means in SAS, skewness and kurtosis were not produced

In terms of distributions of level-1 variables, the unweighted descriptive results from Table 54 suggested that all except for valuing of math, approximated normality, with skewness and kurtosis values within the range of -1.00 and 1.00.

Table 54.

*Unweighted Descriptive Statistics for Level-1 Variables for Egypt (N = 1,876)*

| Variable | *M* | *SD* | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| Overall math achievement | 458.34 | 90.44 | 188.57 | 714.09 | -0.09 | -0.49 |
| Self-confidence in learning math | 1.58 | 0.60 | 0 | 2 | -1.11 | 0.19 |
| Valuing of math | 1.80 | 0.45 | 0 | 2 | -2.14 | 3.88 |
| Time on math homework | 0.92 | 0.63 | 0 | 2 | 0.06 | -0.52 |
| Extra math lessons | 1.59 | 1.01 | 0 | 3 | -0.17 | -1.06 |
| Home resources for learning math | 2.26 | 0.68 | 0 | 3 | -0.53 | -0.22 |

Similarly, a descriptive analysis was conducted on the 19 predictor variables at the school level. Although it appears from Table 55 that, on average, Egyptian students had moderate to high percentage of opportunity to learn math content domains (61.02 for data to 99.13 for number), the range of their minimum opportunity to learn math subjects was quite large, from 16.67 for algebra to 90 for number. Interestingly, in this sample, 100% of math teachers reported being prepared to teach math content. On average, these teachers participated in less than half of the indicated math-related professional development programs (*M* = 2.28, *SD* =1.68) and reported a high level of readiness to teach (*M* = 1.41, *SD* = .52 for data to *M* = 1.96, *SD* = .21 for number).

The data also suggested that in more than half of the lessons, students were given activities related to math instructional practice (*M* = 2.03, *SD* = .19) and math content (*M* = 2.03, *SD* = .13). On average, a moderate amount of homework assignment was assigned to the students (*M* = .96, *SD* = .61). Finally, class size in Egyptian schools tended to have medium class sizes, between 33-40 students (*M* = 1.94, *SD* = .92) and teachers' perception of instructional limitations due to student factors was low (*M* = .23, *SD* = .46). On average, the availability of school resources for math instruction was above medium (*M* = 1.26, *SD* = .76). Noticeably, across 69 schools, the average math instructional hours per years varied greatly, ranging from 22.5 to 174.6, with a mean of 102.06 (*SD* = 44.18).

Also, as shown in Table 55, nine out of 19 level-2 predictor variables appeared to approximate normal distributions with skewness and kurtosis values within the normality approximation range of -1.00 to 1.00. The 11 variables that appeared to depart from normality included opportunity to learn number, algebra, measurement, geometry, preparation to teach, ready to teach number, algebra, measurement, geometry, and teacher perception of math instructional limitations due to student factors.

Table 55.

*Unweighted Descriptive Statistics for Level-2 Variables for Egypt (N =69)*

| Variable | *M* | *SD* | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| Opportunity to learn number | 99.13 | 2.84 | 90 | 100 | -3.00 | 7.19 |
| Opportunity to learn algebra | 91.93 | 13.66 | 16.67 | 100 | -2.86 | 12.54 |
| Opportunity to learn measurement | 91.59 | 14.95 | 25 | 100 | -2.31 | 6.04 |
| Opportunity to learn geometry | 94.56 | 8.85 | 53.85 | 100 | -2.11 | 5.70 |
| Opportunity to learn data | 61.02 | 20.51 | 25 | 100 | 0.39 | -0.63 |
| Amount of homework assignment | 0.96 | 0.65 | 0 | 2 | 0.04 | -0.57 |
| Instructional practice-related activities in math lessons | 2.03 | 0.19 | 1.70 | 2.50 | 0.13 | -0.65 |
| Content-related activities in math lessons | 2.03 | 0.13 | 1.72 | 2.33 | 0.22 | 0.01 |
| Preparation to teach | 1.00 | 0.00 | 1 | 1 | | |
| Ready to teach number | 1.96 | 0.21 | 1 | 2 | -4.58 | 19.52 |
| Ready to teach algebra | 1.94 | 0.24 | 1 | 2 | -3.87 | 13.34 |
| Ready to teach measurement | 1.84 | 0.41 | 0 | 2 | -2.55 | 6.26 |
| Ready to teach geometry | 1.93 | 0.26 | 1 | 2 | -3.37 | 9.65 |
| Ready to teach data | 1.41 | 0.52 | 0 | 2 | 0.07 | -1.30 |
| Math-related professional development | 2.28 | 1.68 | 0 | 5 | 0.16 | -1.09 |
| Class size for math instruction | 1.94 | 0.92 | 0 | 3 | -0.70 | -0.19 |
| School resources for math instruction | 1.26 | 0.76 | 0 | 2 | -0.48 | -1.11 |
| Teacher perception of math instructional limitations due to student factors | 0.23 | 0.46 | 0 | 2 | 1.76 | 2.25 |
| Average math instructional hours per year | 102.06 | 44.18 | 22.5 | 174.6 | -0.42 | -1.12 |

## *Bivariate Analysis*

The results of weighted bivariate correlations among six level-1 variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, extra math lessons, and home resources for learning math) are presented in Appendix I. It appeared from these results that level-1 predictor variables were independent from each other, with correlation coefficients ranging from -.06 between gender and student self-confidence in learning math to .33 between self-confidence in learning math and valuing of math. It was interesting to note that extra math lessons tended to have a negative albeit weak relationship with all level-1 variables ($r$ = -.01 to -.04).

At level-2, unweighted bivariate relationships were estimated for the 19 predictor variables. The correlation matrix for these variables can be found in Appendix K. Unlike level-1, correlation coefficients of level-2 variables had a wider range, from -.30 between class size and ready to teach measurement to .62 between ready to teach number and ready to teach measurement. As expected, correlation coefficients among the variables measuring the same construct tended to be stronger than those measuring different construct. For example, the correlation ranged from .17 to .62 for ready to teach variables. Interestingly, opportunity to learn number had very weak correlations with opportunity to learn data ($r$ = .07). Another interesting relationship was observed between class size and other variables where most of the correlation coefficients were negative in direction.

### *Evaluation of HLM Assumptions*

In order to ensure tenability of results produced by multilevel models in this study, an evaluation of HLM assumptions through visual analysis of both level-1 and level-2 random effects of Model 18 was performed. Model 18 was selected because the results of HLM analysis suggested that it was the most efficient model to predict math achievement in Egypt (see HLM Analysis for Egypt).

The data from Figure 41 suggested that level-1 residuals approximated a normal distribution. In terms of variance, the scatter plot between level-1 residuals and predicted math achievement, as illustrated in Figure 42, suggested that there was evidence of homogeneity of level-1 variance.

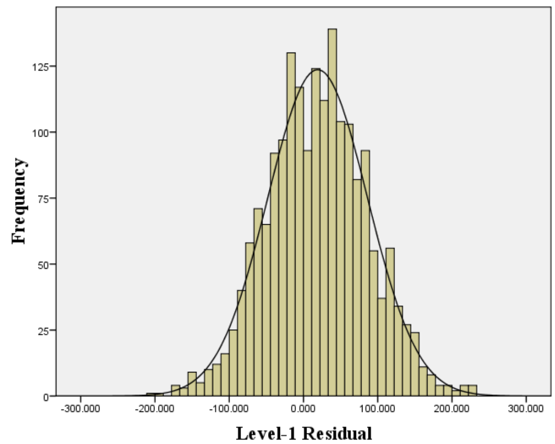



*Figure 41.* Histogram for Level-1 Residuals for Egypt

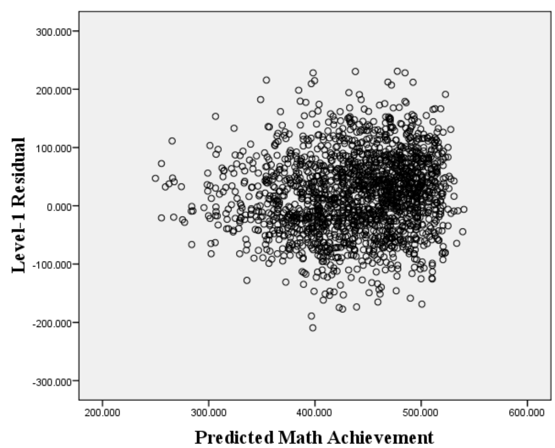



*Figure 42.* Level-1 Residuals by Predicted Math Achievement for Egypt

For level-2 random effects, the empirical Bayes residuals for the intercepts and predicted math scores were used to construct the graphs in Figures 43-44. As can be seen from Figures 43-44, level-2 intercept residuals appeared to have a normal distribution and homogeneous variance.

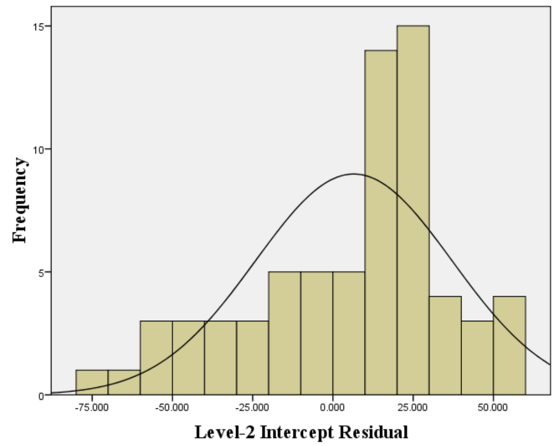



*Figure 43.* Histogram for Level-2 Intercept Residuals for Egypt

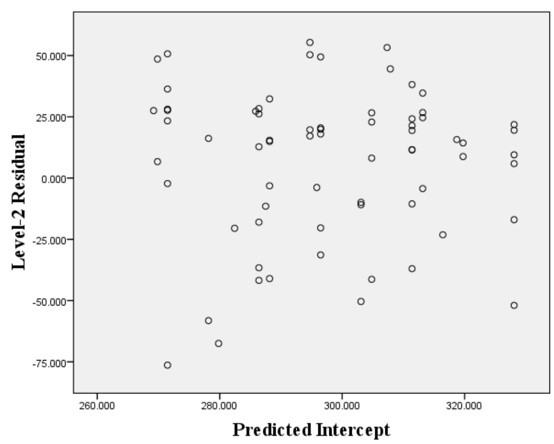



*Figure 44.* Level-2 Intercept Residuals by Predicted Intercept for Egypt

## *HLM Analysis*

### *Unconditional Model (Model 1)*

The HLM analysis started with the unconditional model where none of the level-1 or level-2 predictor was included in the model. The results of the unconditional model are presented in Table 56. For Egypt, the fixed effect for the intercept was 414.26 ($SE$ = 6.38, $p$ <.001). The amount of variability in math achievement was significantly different across schools in Egypt ( $\tau_{00}$ = 2001.88, $SE$ = 44.74, $p$ <.001). Within schools, the amount of unexplained variance was much larger than that between schools ($\sigma^2$ = 5,116.85, *SE* = 71.53). The computed intra-class correlation (ICC) of .28 indicated a modest level of natural clustering of students occurred between schools in Egypt. In other words, approximately 28% of the total variance in math scores occurred between schools.

Table 56.
*Parameter Estimates for Unconditional Model for Egypt*

| Model | Type | Parameters | Estimates | SE | p |
|---|---|---|---|---|---|
| 1 | Fixed | ICC | .28 | | |
| | | INT | 414.26 | 6.38 | <.001 |
| | Random | $\tau_{00}$ | 2001.88 | 44.74 | <.001 |
| | | $\sigma^2$ | 5116.85 | 71.53 | |

*Research Question 1*

To what extent are student background variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, and tutoring in math) associated with TIMSS 2003 eighth-grade math scores in each country?

In order to answer this research question, first, each of the student background variables was entered separately into Model 1 to predict math achievement. Then, as a group of variables, those that contributed significantly in Models 2-6 were included in Model 7 to predict math achievement. Finally, in order to evaluate model fit in terms proportion of variance accounted for, pseudo $R^2$ was computed for the current model against previously constructed models. Results of these models (Models 2-6) are presented in Table 57.

The data from Table 57 suggested that all of the fixed effects estimated by Models 2-6 were statistically significant, except for those of gender in Model 2 ($\gamma$ = -2.86, *SE* = 6.51, *p* = .662) and time students spent on homework in Model 5 ($\gamma$ = 6.24, *SE* = 3.79, *p* = .104). Interestingly, whereas self-confidence in learning math (Model 3) and valuing of math (Model 4) appeared to have statistically significant, positive relationships with math achievement ($\gamma$ = 32.66 and 27.62, *SE* = 3.71 and 6.57; *p* <.001 and .001, respectively); extra math lessons (Model 6) appeared to have a negative relationship with math achievement ($\gamma$ = -6.45, *SE* = 1.81, *p* <.001). Surprisingly, of all slope variances estimated in Models 2-6, only time student spent on homework was statistically significant ($\tau$ = 55.67, *SE* = 7.46, *p* = .028). This suggests that schools in Egypt tended to differ significantly in the relationship between time students spent on homework and math achievement, but were not different in other relationships such as those between math achievement and gender, self-confidence in learning math, student valuing of math, and extra math lessons.

An examination of pseudo $R^2$ across the five models (Models 2-6) suggested that the addition of individual predictors separately to the unconditional model (Model 1) to predict math achievement resulted in reduction of the between school variance in three models (i.e., 26%, 13% and 11% in Models 3, 5, and 6, respectively) and an increase of the between school variance in two models (2% and 39% in Models 2 and 4, respectively). For the within school variance, however, the amount of reduction was smaller, up to 8% in Model 3.

Table 57.
*Parameter Estimates for Models 2-6 (Level-1 Student Background) for Egypt*

| Model | Effect | Parameters | Estimates | *SE* | *p* | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|
| 2 | Fixed | INT | 415.51 | 6.86 | <.001 | | |
| | | Gender | -2.86 | 6.51 | .662 | | |
| | Random | $\tau_{00}$ | 2046.46 | 45.24 | <.001 | | |
| | | Gender | 8.88 | 2.98 | >.500 | | |
| | | $\sigma^2$ | 5118.14 | 71.54 | | | |
| | Pseudo $R^2$ | | | | | -0.02 | 0.00 |

Table 57.

*Parameter Estimates for Models 2-6 (Level-1 Student Background) for Egypt*

| Model | Effect | Parameters | Estimates | *SE* | *p* | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|
| 3 | Fixed | INT | 364.83 | 7.54 | <.001 | | |
| | | Self-confidence | 32.66 | 3.71 | <.001 | | |
| | Random | $\tau_{00}$ | 1481.05 | 38.48 | <.001 | | |
| | | Self-confidence | 119.94 | 10.95 | .193 | | |
| | | $\sigma^2$ | 4706.30 | 68.60 | | | |
| | Pseudo $R^2$ | | | | | 0.26 | 0.08 |
| 4 | Fixed | INT | 365.05 | 13.52 | <.001 | | |
| | | Valuing math | 27.62 | 6.57 | <.001 | | |
| | Random | $\tau_{00}$ | 2775.98 | 52.69 | <.001 | | |
| | | Valuing math | 278.39 | 16.69 | .164 | | |
| | | $\sigma^2$ | 4894.05 | 69.96 | | | |
| | Pseudo $R^2$ | | | | | -0.39 | 0.04 |
| 5 | Fixed | INT | 408.51 | 6.66 | <.001 | | |
| | | Homework time | 6.24 | 3.79 | .104 | | |
| | Random | $\tau_{00}$ | 1749.65 | 41.83 | <.001 | | |
| | | Homework time | 55.67 | 7.46 | .028 | | |
| | | $\sigma^2$ | 5083.99 | 71.30 | | | |
| | Pseudo $R^2$ | | | | | 0.13 | 0.01 |
| 6 | Fixed | INT | 425.33 | 6.58 | <.001 | | |
| | | Extra lessons | -6.45 | 1.81 | .001 | | |
| | Random | $\tau_{00}$ | 1782.77 | 42.22 | <.001 | | |
| | | Extra lessons | 3.05 | 1.75 | >.500 | | |
| | | $\sigma^2$ | 5081.68 | 71.29 | | | |
| | Pseudo $R^2$ | | | | | 0.11 | 0.01 |

Note: Pseudo $R^2$ refers to the difference in proportion of variance accounted for between the current models (Models 2-6) and the unconditional model (Model 1).

As a next step of model building, all of the student background variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, and tutoring in math) were included in the combined model, Model 7, to predict math achievement. Notably, in the presence of other level-1 variables in the model, all but gender had statistically significant fixed effects. With a fixed effect of -5.90 (*SE* = 1.68, *p* <.001) for extra math lesson, it could be inferred that for each unit increase in extra math lesson (i.e., from 0 for never to 3 for daily), the students were expected to reduce 5.90 points in their math scores while controlling for other predictors in the model. Similarly, with fixed effect of 29.80 (*SE* = 3.36, *p* <.001) for self-confidence in learning math, it could be interpreted that for each unit increase in level of self-confidence in learning math (i.e., from 0 for low to 2 for high), it was expected that the students would improve 29.80 points in their math scores while controlling for other predictors in the model. Likewise, each unit change in student valuing of math was associated with an increase of 14.22 points (*SE* = 6.52, *p* = .033) in math achievement and each unit change in time student spent on homework was associated with an increase of 7.77 points (*SE* = 3.26, *p* = .020), an after adjusting for other predictors in the model.

Interestingly, it was noted that all of the random effects in this model were not statistically significant (p > .50). This indicates that the observed relationships between math achievement and student background variables did not seem to vary significantly across schools in Egypt.

Table 58.
*Parameter Estimates for Model 7 (Level-1 Student Background) for Egypt*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 7 | Fixed | INT | 349.43 | 13.66 | <.001 |
| | | Gender | -5.71 | 5.24 | .280 |
| | | Extra lessons | -5.90 | 1.68 | .001 |
| | | Self-confidence | 29.80 | 3.36 | <.001 |
| | | Valuing math | 14.22 | 6.52 | .033 |
| | | Homework time | 7.77 | 3.26 | .020 |
| | Random | $\tau_{00}$ | 2887.49 | 53.74 | >.500 |
| | | Gender | 4.80 | 2.19 | >.500 |
| | | Extra lessons | 8.29 | 2.88 | >.500 |
| | | Self-confidence | 93.49 | 9.67 | >.500 |
| | | Valuing math | 333.02 | 18.25 | >.500 |
| | | Homework time | 60.17 | 7.76 | >.500 |

Table 58.
*Parameter Estimates for Model 7 (Level-1 Student Background) for Egypt*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | $\sigma^2$ | 4500.18 | 67.08 | |

An evaluation of model fit was also conducted between Model 7 and previously constructed models, Models 2-6. As can be seen from Table 59, the inclusion of student background variables in Model 7 resulted in some reduction in the amount of variance accounted for in math achievement within schools (from 4% when compared with Model 3 and 12% when compared with Model 2). Unexpectedly, between schools, the amount of variance appeared to increase notably, from 4% in Model 4 to 95% in Model 3. In sum, Model 7 was more efficient than earlier models in that it accounted for more variance in math achievement within schools in Egypt. However, this model appeared to be less efficient than previously constructed models in that it accounted for less variance in math achievement between schools in Egypt.

Table 59.
*Comparison of $R^2$ between Model 7 and Previously Constructed Models for Egypt*

| Previous Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 2 | -0.41 | 0.12 |
| 3 | -0.95 | 0.04 |
| 4 | -0.04 | 0.08 |
| 5 | -0.65 | 0.11 |
| 6 | -0.62 | 0.11 |

*Research Question 2*

To what extent are home resources variables (i.e., availability of calculator, computer, and desk for student use) associated with TIMSS 2003 eighth-grade math scores in each country?

When the level-1 predictor home resources was added to the unconditional model to predict math achievement, a reduction of 5% in the within school variance and an increase of 1% in the between school variance was noted (see Table 60). As a fixed effect, home resources was statistically significant ($\gamma$ = 31.92, *SE* = 6.8, *p* <.001), meaning that each unit change in home resources (i.e., from 0 to 3) was associated with an increase of 31.92 points in student math achievement, while not controlling for other variables in the model. As a random effect, home resources was not statistically significant (τ = 10.69, *SE* = 3.27, *p* = 3.54) which indicates that the relationship between home resources and math achievement did not vary significantly across schools in Egypt.

Table 60.
*Parameter Estimates for Level-1 Home Resources Model for Egypt*

| Model | Effect | Parameters | Estimates | SE | P | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|
| 8 | Fixed | INT | 352.86 | 9.37 | <.001 | | |
| | | Home resources | 31.92 | 3.65 | <.001 | | |
| | Random | $\tau_{00}$ | 2011.95 | 44.85 | <.001 | | |
| | | Home resources | 10.69 | 3.27 | .354 | | |
| | | $\sigma^2$ | 4853.60 | 69.67 | | | |
| | Pseudo $R^2$ | | | | | -0.01 | 0.05 |

Note: Pseudo $R^2$ refers to the difference in the proportion variance between Model 8 and Model 1.

Given the findings obtained from Models 7 and 8, five out of six student-related variables were entered into the unconditional model to make Model 9. Gender was excluded from Model 9 because both of its fixed and random effects were not statistically significant in Model 7. Also, in Model 9, all the slope variances were set to 0 because they were not statistically significant in earlier models.

As can be seen from Table 61, all of the level-1 variables had statistically significant fixed effects. Of the five variables in the models, home resources appears to have the strongest positive relationship with math achievement ($\gamma$ = 28.83, *SE* = 3.74, *p* <.001). Following was student self-confidence in learning math with a fixed effect of 26.94 (*SE* = 3.30, *p* <.001). Next were student valuing of math ($\gamma$ = 15.94, *SE* = 8.00, *p*

=.046) and time student spent on homework ($\gamma$ = 7.04, *SE* = 2.88, *p* = .015). As for extra math lesson, an inverse relationship was noted between this predictor and math achievement ($\gamma$ = -5.64, *SE* = 1.75, *p* = .002). These results suggest that the more resources students had at home, the more self-confidence students expressed in learning math, the higher students valued math, and the more time students spent on their homework, the higher math scores the students tended to achieve. In contrast, it appears that the more frequently students took extra math lessons, the poorer math scores they seem to earn.

In terms of random effects, the amount of 1,382.87 (*SE* = 37.19, *p* <.001) for the between school variance was statistically significant, suggesting that a considerable amount of variance between schools remained to be explained.

As compared to Models 7 and 8, Model 9 appeared more efficient because it accounted for a significantly higher amount of the explained variance between schools, up to 52% when compared to Model 7. In terms of the explained variance within schools, an increase of up to 8% was observed. As a result of these comparisons, Model 9 was selected as the foundational level-1 model for further examination of the relationships between level-2 predictors and math achievement.

Table 61.
*Parameter Estimates for Combined Level-1 Predictors Model for Egypt*

| Model | Type | Parameters | Estimates | *SE* | *p* | Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|
| 9 | Fixed | INT | 292.77 | 17.45 | <.001 | | | |
| | | Extra lessons | -5.64 | 1.75 | .002 | | | |
| | | Self-confidence | 26.94 | 3.30 | <.001 | | | |
| | | Valuing math | 15.94 | 8.00 | .046 | | | |
| | | Homework time | 7.04 | 2.88 | .015 | | | |
| | | Home resources | 28.83 | 3.74 | <.001 | | | |
| | Random | $\tau_{00}$ | 1382.87 | 37.19 | <.001 | | | |
| | | $\sigma^2$ | 4442.74 | 66.65 | | | | |

Table 61.
*Parameter Estimates for Combined Level-1 Predictors Model for Egypt*

| Model | Type | Parameters | Estimates | *SE* | *p* | Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|
| | Pseudo $R^2$ | | | | | 7 | 0.52 | 0.01 |
| | | | | | | 8 | 0.31 | 0.08 |

Note: Pseudo $R^2$ refers to the difference in the proportion of variance between Model 9 and Models 7-8.

*Research Question 3*

To what extent are instructional variables (i.e., opportunity to learn, activities in math lessons, amount of homework assignment, and instructional time) associated with TIMSS 2003 eighth-grade math scores in each country?

In addressing this research question, a similar strategy for model building used in Research Question 1 was applied here. That is, each of the level-2 instructional practice variables was first added to the foundational level-1 model (Model 9) to make Models 10-13. Then, as a group, those variables with significant fixed effects in Models 10-13 were included in the combined instructional practices model, Model 14. It is important to note that in these models, there was no cross-level interactions between level-1 and level-2 predictors because, as described earlier in Model 9, all of the random slopes were set to 0. The results of Models 10-14 are presented in Tables 62-65.

As can be seen in Table 62, none of the level-2 predictors that were separately added to Models 10-13 was statistically significant. This means that, statistically, there was no evidence that these instructional practices-related variables contributed significantly to predict math achievement across schools in Egypt. Interestingly, however, in the presence of these level-2 variables in the models, all the level-1 predictors were found statistically significant and the sizes of their fixed effects were similar to what were observed in Model 9.

Table 62.

*Parameter Estimates for Level-2 Instructional Practices Models for Egypt*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 10 | Fixed | INT | 359.15 | 131.28 | .009 |
| | | Opportunity_algebra | -0.28 | 0.34 | .411 |
| | | Opportunity_data | 0.13 | 0.27 | .623 |
| | | Opportunity_geometry | -0.43 | 0.64 | .507 |
| | | Opportunity_measurement | 0.78 | 0.39 | .051 |
| | | Opportunity_number | -0.78 | 1.38 | .573 |
| | | Extra lessons | -5.65 | 1.74 | .002 |
| | | Self-confidence | 26.94 | 3.31 | <.001 |
| | | Valuing math | 15.82 | 7.88 | .044 |
| | | Homework time | 7.09 | 2.88 | .014 |
| | | Home resources | 28.76 | 3.69 | <.001 |
| | Random | $\tau_{00}$ | 1301.23 | 36.07 | <.001 |
| | | $\sigma^2$ | 4442.80 | 66.65 | |
| 11 | Fixed | INT | 283.86 | 18.14 | <.001 |
| | | Homework assignment | 11.20 | 6.27 | .078 |
| | | Extra lessons | -5.66 | 1.75 | .002 |
| | | Self-confidence | 26.94 | 3.31 | <.001 |
| | | Valuing math | 15.86 | 7.96 | .046 |
| | | Homework time | 7.10 | 2.88 | .014 |
| | | Home resources | 28.69 | 3.73 | <.001 |
| | Random | $\tau_{00}$ | 1355.07 | 36.81 | <.001 |
| | | $\sigma^2$ | 4442.49 | 66.65 | |
| 12 | Fixed | INT | 445.37 | 98.84 | <.001 |
| | | Content_activities | -72.97 | 39.91 | .072 |
| | | Instruction_activities | -2.72 | 29.49 | .927 |
| | | Extra lessons | -5.58 | 1.74 | .002 |
| | | Self-confidence | 26.87 | 3.29 | <.001 |
| | | Valuing math | 16.01 | 7.99 | .045 |
| | | Homework time | 6.94 | 2.87 | .016 |
| | | Home resources | 28.94 | 3.76 | <.001 |
| | Random | $\tau_{00}$ | 1346.18 | 36.69 | <.001 |
| | | $\sigma^2$ | 4443.11 | 66.66 | |
| 13 | Fixed | INT | 292.65 | 17.37 | <.001 |
| | | Instructional hours | -0.03 | 0.14 | .840 |
| | | Extra lessons | -5.63 | 1.75 | .002 |
| | | Self-confidence | 26.95 | 3.31 | <.001 |
| | | Valuing math | 15.95 | 7.99 | .046 |
| | | Homework time | 7.03 | 2.87 | .015 |
| | | Home resources | 28.83 | 3.73 | <.001 |
| | Random | $\tau_{00}$ | 1403.31 | 37.46 | <.001 |
| | | $\sigma^2$ | 4442.88 | 66.65 | |

In terms of model fit, in comparison with the foundational level-1 model (Model 9), Model 10 appeared to be the most efficient model because the amount of explained variance between schools in this model increased by 6% (see Table 63). As for the within school variance, no significant difference was observed between Models 10-13 and Model 9 (pseudo $R^2$ = 0).

Table 63.
*Comparison of $R^2$ between Level-2 Instructional Practice Models and Foundational Level-1 Model for Egypt*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 10 vs. 9 | 0.06 | 0.00 |
| 11 vs. 9 | 0.02 | 0.00 |
| 12 vs. 9 | 0.03 | 0.00 |
| 13 vs. 9 | -0.01 | 0.00 |

Similar to what was observed in Models 10-13, when using all the level-2 instructional practice variables to predict math achievement, Model 14 did not produce any statistically significant level-2 fixed effects (see Table 64). These results suggest that, in Egypt, instructional practices-related predictors did not appear to contribute significantly in the statistical prediction of student math achievement. However, with statistically significant level-1 fixed effects and random effects, it could be inferred that extra math lessons, student self-confidence in learning math, student valuing of math, time students spent on homework, and home resources tended to be good predictors of math achievement and that schools in Egypt varied significantly in their mean math performance.

Table 64.
*Parameter Estimates for The Combined Level-2 Instructional Practices Model for Egypt*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 14 | Fixed | INT | 436.27 | 165.51 | .011 |
| | | Homework assignment | 3.99 | 7.53 | .598 |
| | | Opportunity_algebra | -0.24 | 0.38 | .523 |
| | | Opportunity_data | 0.15 | 0.27 | .571 |

Table 64.
*Parameter Estimates for The Combined Level-2 Instructional Practices Model for Egypt*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Opportunity_geometry | -0.50 | 0.64 | .439 |
| | | Opportunity_measurement | 0.78 | 0.43 | .074 |
| | | Opportunity_number | 0.61 | 1.39 | .664 |
| | | Content_activities | -68.38 | 43.76 | .123 |
| | | Instruction_activities | 21.03 | 30.94 | .499 |
| | | Instructional hours | -0.16 | 0.13 | .219 |
| | | Extra lessons | -5.61 | 1.75 | .002 |
| | | Self-confidence | 26.89 | 3.31 | <.001 |
| | | Valuing math | 15.83 | 7.92 | .045 |
| | | Homework time | 7.02 | 2.86 | .014 |
| | | Home resources | 28.79 | 3.75 | <.001 |
| | Random | $\tau_{00}$ | 1294.75 | 35.98 | <.001 |
| | | $\sigma^2$ | 4443.16 | 66.66 | |

As evident in Table 65, compared to previously constructed models (Models 9-13), Model 14 appears to be more efficient because this model accounted for more variance between schools. For example, an increase of 8% in the between school variance was observed when using Model 14 instead of Models 13. However, there was one exception. That is, changing from Models 10 to Model 14 would result in no difference in the amount of explained variance accounted for within and between schools. Therefore, in consideration of the amount of the explained variance between schools, Model 10 served as the most efficient and parsimonious instructional practices-related model in predicting math achievement in Egypt.

Table 65.
*Comparison of $R^2$ between Model 14 and Previously Constructed Models 9-13 for Egypt*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 14 vs. 9 | 0.06 | 0.00 |
| 14 vs. 10 | 0.00 | 0.00 |
| 14 vs. 11 | 0.04 | 0.00 |
| 14 vs. 12 | 0.04 | 0.00 |
| 14 vs. 13 | 0.08 | 0.00 |

*Research Question 4*

To what extent are teacher-related variables (i.e., preparation to teach, ready to teach, and professional development) associated with TIMSS 2003 eighth-grade math scores in each country?

Similarly, incremental model building strategies were applied to examine the relationships among teacher-related variables (i.e., preparation to teach, ready to teach, and professional development) and math achievement. It is worth noting, however, that because the predictor preparation to teach math content had no variation in Egypt (i.e., 100% math teachers in this sample reported being prepared to teach), Model 15 could not be estimated. Thus, only results of Models 16-18 are presented in Tables 66-68.

As shown in Table 66, of all the level-2 predictors, only math-related professional development in Model 16 had a statistically significant fixed effect ($\gamma$ = 8.32, *SE* = 3.12, *p* = .010). This means each unit increases in math-related professional development was associated with an increase of 8.32 points in student math scores, after adjusting for other variables in the models. In terms of level-1 fixed effects and random effects, all in Models 16 and 17 were statistically significant. These results suggested that a significant amount of variance in math achievement remained unexplained between schools in Egypt.

When comparing the proportion of variance accounted for by Models 16 and 17 with that of the foundational level-1 model (Model 9), it appears that Model 16 was the most efficient with the amount of variance between school increased by 10% (see Table 66). No improvement in the within school variance was noted by use of these models.

Table 66.

*Parameter Estimates for Teacher Background Models for Egypt*

| Model | Type | Parameters | Estimates | *SE* | *p* | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|
| 16 | Fixed | INT | 356.35 | 53.41 | <.001 | | |
| | | Ready_number | -32.27 | 36.29 | .378 | | |
| | | Ready_algebra | -21.98 | 23.94 | .363 | | |
| | | Ready_Measurement | 14.16 | 15.58 | .367 | | |
| | | Ready_Geometry | -1.38 | 11.71 | .907 | | |
| | | Ready_Data | 15.11 | 12.62 | .236 | | |
| | | Extra lessons | -5.66 | 1.76 | .002 | | |
| | | Self-confidence | 26.87 | 3.31 | <.001 | | |
| | | Valuing math | 15.99 | 7.94 | .044 | | |
| | | Homework time | 6.86 | 2.87 | .017 | | |
| | | Home resources | 28.62 | 3.68 | <.001 | | |
| | Random | $\tau_{00}$ | 1372.70 | 37.05 | <.001 | | |
| | | $\sigma^2$ | 4442.89 | 66.65 | | | |
| | Pseudo $R^2$ | | | | | 0.01 | 0.00 |
| 17 | Fixed | INT | 276.93 | 18.27 | <.001 | | |
| | | Professional development | 8.32 | 3.12 | .010 | | |
| | | Extra lessons | -5.60 | 1.73 | .002 | | |
| | | Self-confidence | 27.07 | 3.29 | <.001 | | |
| | | Valuing math | 16.02 | 7.95 | .044 | | |
| | | Homework time | 6.97 | 2.86 | .015 | | |
| | | Home resources | 28.72 | 3.70 | <.001 | | |
| | Random | $\tau_{00}$ | 1238.67 | 35.19 | <.001 | | |
| | | $\sigma^2$ | 4442.64 | 66.65 | | | |
| | Pseudo $R^2$ | | | | | 0.10 | 0.00 |

Note:
Model 15 could not be computed because there was no variation for the level-2 predictor Preparation to teach math content.
Pseudo $R^2$ refers to the difference in the proportion of variance between Model 9 and Models 16-17.

In Model 18 where ready to teach math topics and math-related professional development were included as level-2 predictors to predict math achievement, only math-related professional development was found statistically significant ($\gamma$ = 8.34, *SE* = 3.37, *p* = .016). This suggests that the more math-related professional development programs that Egyptian teachers took, the higher math scores their students tended to achieve. Specifically, with every unit increases in math-related professional development, student math scores were expected to increase by 8.34 points, after adjusting for other variables in the model. Also, all the level-1 fixed effects and all the random effects were statistically significant, meaning that, in this model, extra math lessons, student self-confidence in learning math, student valuing of math, time students spent on homework, and home resources were good predictors of math achievement, and that there was a significant amount of variability in average math achievement across schools in Egypt.

Table 67.

*Parameter Estimates for the Combined Teacher Background Model for Egypt*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 18 | Fixed | INT | 349.77 | 55.37 | <.001 |
| | | Professional development | 8.34 | 3.37 | .016 |
| | | Ready_number | -29.70 | 36.92 | .424 |
| | | Ready_algebra | -27.53 | 25.18 | .279 |
| | | Ready_measurement | 9.98 | 16.64 | .551 |
| | | Ready_geometry | 0.64 | 12.41 | .960 |
| | | Ready_data | 14.93 | 11.58 | .203 |
| | | Extra lessons | -5.64 | 1.75 | .002 |
| | | Self-confidence | 26.99 | 3.31 | <.001 |
| | | Valuing math | 16.10 | 7.92 | .042 |
| | | Homework time | 6.76 | 2.86 | .018 |
| | | Home resources | 28.52 | 3.66 | <.001 |
| | Random | $\tau_{00}$ | 1224.62 | 34.99 | <.001 |
| | | $\sigma^2$ | 4442.82 | 66.65 | |

As evident in Table 68, Model 18 appeared to be more efficient than Models 9, 16 and 17 in terms of the amount of variance accounted for between schools. Specifically, an increase of 1% to 11% in the between school variance was likely to result when using Model 18 as opposed to Models 9, 16, or 17.

Table 68.

*Comparison of $R^2$ between Model 18 and Previously Constructed Models 9 and 16-17 for Egypt*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 18 vs. 9 | 0.11 | 0.00 |
| 18 vs. 16 | 0.11 | 0.00 |
| 18 vs. 17 | 0.01 | 0.00 |

*Research Question 5*

To what extent are school-related variables (i.e., class size, school resources for math instruction, and teacher perception of math instructional limitations due to student factors) associated with TIMSS 2003 eighth-grade math scores in each country?

Table 69 provides a summary of the results for Models 19-21 where school-related variables (i.e., class size, school resources for math instruction, and teacher perception of math instructional limitations due to student factors) were separately included in the models to predict math achievement. In these models, no statistically significant level-2 main effects were detected, meaning these predictors did not appear to predict well math achievement in Egypt. Again, similar to what was observed in earlier models, all of the level-1 fixed and random effects were statistically significant.

Table 69.
*Parameter Estimates for School Background Models for Egypt*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 19 | Fixed | INT | 309.31 | 22.68 | <.001 |
| | | Class size | -7.63 | 7.90 | .338 |
| | | Extra lessons | -5.62 | 1.75 | .002 |
| | | Self-confidence | 27.03 | 3.32 | <.001 |
| | | Valuing math | 16.03 | 8.01 | .045 |
| | | Homework time | 7.00 | 2.89 | .016 |
| | | Home resources | 28.80 | 3.73 | <.001 |
| | Random | $\tau_{00}$ | 1355.58 | 36.82 | <.001 |
| | | $\sigma^2$ | 4443.74 | 66.66 | |
| 20 | Fixed | INT | 291.41 | 17.84 | <.001 |
| | | Instructional limitation | 5.84 | 7.40 | .433 |
| | | Extra lessons | -5.64 | 1.76 | .002 |
| | | Self-confidence | 26.92 | 3.30 | <.001 |
| | | Valuing math | 15.90 | 7.99 | .046 |
| | | Homework time | 7.05 | 2.89 | .015 |
| | | Home resources | 28.81 | 3.74 | <.001 |
| | Random | $\tau_{00}$ | 1398.20 | 37.39 | <.001 |
| | | $\sigma^2$ | 4442.66 | 66.65 | |
| 21 | Fixed | INT | 289.66 | 20.39 | <.001 |
| | | School resources | 2.92 | 7.24 | .687 |

Table 69.
*Parameter Estimates for School Background Models for Egypt*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Extra lessons | -5.64 | 1.75 | .002 |
| | | Self-confidence | 26.93 | 3.30 | <.001 |
| | | Valuing math | 15.90 | 7.97 | .046 |
| | | Homework time | 7.07 | 2.87 | .014 |
| | | Home resources | 28.81 | 3.73 | <.001 |
| | Random | $\tau_{00}$ | 1402.07 | 37.44 | <.001 |
| | | $\sigma^2$ | 4442.63 | 66.65 | |

In comparing Models 19-21 with Model 9 in terms of the proportion of variance accounted for, it looks like that Model 19 was slightly more efficient (see Table 70). Specifically, the use of Model 19 as opposed to Model 9 increased the amount of between school variance accounted for by 2%; whereas the use of Models 20 and 21 as opposed to Model 9 accounted resulted in an increase of 1% in the between school variance.

Table 70.
*Comparison of $R^2$ between Level-2 Teacher Background and Foundational Level-1 Model for Egypt*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 19 vs. 9 | 0.02 | 0.00 |
| 20 vs. 9 | -0.01 | 0.00 |
| 21 vs. 9 | -0.01 | 0.00 |

As shown in Table 71, Model 22 with all of the school background-related predictors included in the model to predict math achievement did not produce any statistically significant level-2 main effects. These results suggest that, after controlling for all other level-1 and level-2 variables in the model, students in schools with larger class sizes did not statistically perform better in math than their peers in schools with smaller class sizes. Similarly, it seems that there was no statistically significant difference in math achievement among students in schools where teachers perceived to have more limitations due to student factors and those in schools where teachers perceived to have none or few limitations due to student factors. Likewise, school resources did not show significant relationship with math achievement across schools in Egypt, after adjusting for other predictors in the model.

As expected, all the level-1 fixed effects and random effects in Model 22 were found statistically significant. This indicated that the amount of within and between school variance that remained to be explained was still significant.

Table 71.
*Parameter Estimates for the Combined School Background Model for Egypt*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 22 | Fixed | INT | 306.86 | 24.54 | <.001 |
| | | Instructional limitation | 6.78 | 8.39 | .422 |
| | | Class size | -8.19 | 7.89 | .304 |
| | | School resources | 1.97 | 7.46 | .793 |
| | | Extra lessons | -5.62 | 1.77 | .002 |
| | | Self-confidence | 26.99 | 3.33 | <.001 |
| | | Valuing math | 15.97 | 7.97 | .045 |
| | | Homework time | 7.02 | 2.88 | .015 |
| | | Home resources | 28.75 | 3.73 | <.001 |
| | Random | $\tau_{00}$ | 1387.90 | 37.25 | <.001 |
| | | $\sigma^2$ | 4443.52 | 66.66 | |

When comparing Model 22 against earlier constructed models, it appears that this model was not the most efficient (see Table 72). Although the amount of explained variance between schools in Model 22 was better than Models 20 and 21 and equal to Model 9, it accounted for less variance between schools than Model 19 by 2%. Thus, Model 19 served as the most efficient school-related model to predict math achievement in Egypt.

Table 72.
*Comparison of $R^2$ between Model 22 and Previously Constructed Models 9 and 19-21 for Egypt*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 22 vs. 9 | 0.00 | 0.00 |
| 22 vs. 19 | -0.02 | 0.00 |
| 22 vs. 20 | 0.01 | 0.00 |
| 22 vs. 21 | 0.01 | 0.00 |

*Final Model*

With an intention to identify the most efficient and parsimonious model to predict eighth-grade math achievement in Egypt, Model 23 was built by including all the statistically significant level-2 predictors in earlier combined models to Model 9 and then compared with the three combined models, Models 14, 18, and 22.

As can be seen from Table 73, all of the fixed and random effects estimated in Model 23 were statistically significant. Of the fixed effects, five showed positive relationships with math achievement and one showed a negative relationship with math achievement. As an example, whereas an increase in home resources was associated with an improvement of 28.72 points in student math achievement, a unit change in extra math lessons was associated with a reduction in student math scores by 5.60 points, while controlling for other variables. As for math-related professional development, for every extra program Egyptian teachers took, it was likely that their students would increase their math scores by 8.32 points, while controlling for other variables in the model. Likewise, the more time students spent on homework and the higher level of self-confidence in learning math and valuing of math that students expressed, the higher math scores they tended to achieve. In addition, it seems that significant variation existed in math achievement across schools in Egypt.

Table 73.
*Parameter Estimates for Full Model for Egypt*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 23 | Fixed | INT | 276.93 | 18.27 | <.001 |
| | | Professional development | 8.32 | 3.12 | .010 |
| | | Extra lessons | -5.60 | 1.73 | .002 |
| | | Self-confidence | 27.07 | 3.29 | <.001 |
| | | Valuing math | 16.02 | 7.95 | .044 |
| | | Homework time | 6.97 | 2.86 | .015 |
| | | Home resources | 28.72 | 3.70 | <.001 |
| | Random | $\tau_{00}$ | 1238.67 | 35.19 | <.001 |

Table 73.
*Parameter Estimates for Full Model for Egypt*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | $\sigma^2$ | 4442.64 | 66.65 | |

Finally, as shown in Table 74, the amount of between school variance accounted for by Model 23 was 4% larger than that accounted for by Model 14 and 11% larger than that accounted for by Model 22. However, in comparison with Model 18, Model 23 accounted for less variance between schools by 1%. Thus, in consideration of these results, Model 18 served as the best model for predicting math achievement in Egypt.

Table 74.
*Comparison of $R^2$ between Model 23 and Previously Constructed Models 14, 18 and 22 for Egypt*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 14 | 0.04 | 0.00 |
| 18 | -0.01 | 0.00 |
| 22 | 0.11 | 0.00 |

## Results for South Africa

### *Evaluation of Missing Data*

As a result of the listwise deletion method, the sample size for South Africa was reduced from 8,927 students and 253 schools to 1,564 students and 52 schools. This means only 17.52% of the original sample had complete data on all variables of interest in this study. In order to evaluate the extent to which the data for South Africa were missing completely at random, the missingness on 19 level-2 variables was correlated. Results of this analysis suggested a non-randomness of missing data, with the majority of the missingness indicators (16) having moderate to strong correlation coefficients with each other (*r* ranged from .50 to .99). In addition, when missingness was correlated with values of itself as well as values of other variables, only marginal correlations were observed ($r$ = -.27 to .18). In summary, the missing data mechanism for South Africa was not missing completely at random.

### *Univariate Analysis*

A descriptive examination of level-1 variables (i.e., overall math achievement, gender, self-confidence in learning math, valuing of math, time on math homework, extra math lessons, and home resources for learning math) was conducted using SAS 9.13. Of the complete sample of 1,564 eighth-grade students, 773 (49.42%) were male and 791 (50.58%) were female. On average, the weighted overall math achievement for South African students was relatively low 272.66 (*SD* = 108.11) with the lowest score of 70.60 and the highest score of 670.84 (see Table 75).

With regard to level-1 predictor variables, it appeared that, on average, eighth-grade students in South Africa had moderate support at home in terms of resources for learning (*M* = 1.66 , *SD* = .89), were above medium level of self-confidence in learning math (*M* = 1.23, *SD* = .67) and valuing of math (*M* = 1.77, *SD* = .49), spent a modest amount of time on math homework (*M* = .99, *SD* = .64), and took extra math lessons about one to two times a week (*M* = 1.49, *SD* = 1.13) (see Table 75).

Table 75.

*Weighted Descriptive Statistics for Level-1 Variables for South Africa (N = 1,564)*

| Variable | *M* | *SD* | Min | Max |
|---|---|---|---|---|
| Overall math achievement | 272.66 | 108.11 | 70.60 | 670.84 |
| Self-confidence in learning math | 1.23 | 0.67 | 0.00 | 2.00 |
| Valuing of math | 1.77 | 0.49 | 0.00 | 2.00 |
| Time on math homework | 0.99 | 0.64 | 0.00 | 2.00 |
| Extra math lessons | 1.49 | 1.13 | 0.00 | 3.00 |
| Home resources for learning math | 1.66 | 0.89 | 0.00 | 3.00 |

Note: When weight was used to compute means in SAS, skewness and kurtosis were not produced

In terms of distributions of level-1 variables, the unweighted descriptive results from Table 76 suggested that all except for valuing of math, approximated normality, with skewness and kurtosis values within the range of -1.00 and 1.00.

Table 76.

*Unweighted Descriptive Statistics for Level-1 Variables for South Africa (N = 1,564)*

| Variable | *M* | *SD* | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| Overall math achievement | 268.54 | 94.31 | 70.60 | 670.84 | 1.16 | 1.67 |
| Self-confidence in learning math | 1.23 | 0.66 | 0.00 | 2.00 | -0.30 | -0.77 |
| Valuing of math | 1.78 | 0.48 | 0.00 | 2.00 | -2.15 | 3.91 |
| Time on math homework | 1.00 | 0.64 | 0.00 | 2.00 | 0.00 | -0.58 |
| Extra math lessons | 1.54 | 1.11 | 0.00 | 3.00 | 0.08 | -1.35 |
| Home resources for learning math | 1.63 | 0.88 | 0.00 | 3.00 | -0.13 | -0.68 |

Similarly, a descriptive analysis was conducted on 19 predictor variables at the school level. As shown in Table 77, on average, South African students had modest to moderate percentage of opportunity to learn math content domains, from 39.32% (*SD* = 34.32) for data to 75.30% (*SD* = 24.16) for number. Although not every math teacher reported being prepared to teach math content (*M* = .75, *SD* = .44), on average, they participated in various types of math-related professional development programs (*M* = 3.04, *SD* =1.80) and reported a relatively high level of readiness to teach (*M* = 1.37, *SD* = .66 for data to *M* = 1.88, *SD* = .32 for number).

The data also suggested that in about half of the lessons, students were given activities related to math instructional practice (*M* = 2.18, *SD* = .23) and math content (*M* = 1.94, *SD* = .23). On average, a moderate amount of homework assignment was assigned to the students (*M* = 1.10, *SD* = .72). Finally, South African schools tended to have relatively large class sizes, more than 40 students (*M* = 2.42, *SD* = .85) and teachers' perception of instructional limitations due to student factors was close to medium (*M* = .90, *SD* = .75). On average, the availability of school resources for math instruction was relatively low (*M* = .77, *SD* = .65). Noticeably, across 52 schools, the average math instructional hours per year varied greatly, from 93 to 360, with a mean of 164.14 (*SD* = 57.11).

Also, as shown in Table 77, the majority of level-2 predictor variables (16 out of 19) appeared to approximate normal distributions with skewness and kurtosis values within the normality approximation range of -1.00 to 1.00. The three variables that appeared to depart from normality included content-related activities in math lessons, ready to teach number, and average math instructional hours per year.

Table 77.

*Unweighted Descriptive Statistics for Level-2 Variables for South Africa (N =52)*

| Variable | *M* | *SD* | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| Opportunity to learn number | 75.30 | 24.16 | 0.00 | 100.00 | -0.86 | 0.34 |
| Opportunity to learn algebra | 57.60 | 30.08 | 0.00 | 100.00 | -0.09 | -1.03 |
| Opportunity to learn measurement | 50.34 | 32.29 | 0.00 | 100.00 | -0.22 | -1.03 |
| Opportunity to learn geometry | 47.72 | 27.26 | 0.00 | 100.00 | 0.61 | -0.21 |
| Opportunity to learn data | 39.32 | 34.72 | 0.00 | 100.00 | 0.36 | -1.04 |
| Amount of homework assignment | 1.10 | 0.72 | 0.00 | 2.00 | -0.15 | -1.02 |

Table 77.
*Unweighted Descriptive Statistics for Level-2 Variables for South Africa (N =52)*

| Variable | *M* | *SD* | Min | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| Instructional practice-related activities in math lessons | 2.18 | 0.23 | 1.73 | 2.65 | -0.16 | -0.82 |
| Content-related activities in math lessons | 1.94 | 0.23 | 1.21 | 2.22 | -1.47 | 2.21 |
| Preparation to teach | 0.75 | 0.44 | 0.00 | 1.00 | -1.19 | -0.61 |
| Ready to teach number | 1.88 | 0.32 | 1.00 | 2.00 | -2.48 | 4.31 |
| Ready to teach algebra | 1.65 | 0.52 | 0.00 | 2.00 | -1.10 | 0.12 |
| Ready to teach measurement | 1.44 | 0.67 | 0.00 | 2.00 | -0.80 | -0.42 |
| Ready to teach geometry | 1.62 | 0.60 | 0.00 | 2.00 | -1.32 | 0.79 |
| Ready to teach data | 1.37 | 0.66 | 0.00 | 2.00 | -0.55 | -0.63 |
| Math-related professional development | 3.04 | 1.80 | 0.00 | 5.00 | -0.37 | -1.33 |
| Class size for math instruction | 2.42 | 0.85 | 0.00 | 3.00 | -1.35 | 0.97 |
| School resources for math instruction | 0.77 | 0.65 | 0.00 | 2.00 | 0.25 | -0.62 |
| Teacher perception of math instructional limitations due to student factors | 0.90 | 0.75 | 0.00 | 2.00 | 0.16 | -1.16 |
| Average math instructional hours per year | 164.14 | 57.11 | 93.00 | 360.00 | 1.55 | 2.36 |

## *Bivariate Analysis*

The results of weighted bivariate correlations among six level-1 variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, extra math lessons, and home resources for learning math) are presented in Appendix L. It appeared from these results that level-1 predictor variables were uncorrelated with each other, with correlation coefficients ranging from -.07 between gender and student self-confidence in learning math to .18 between self-confidence in learning math and valuing of math. It was interesting to note that extra math lessons tended to have a negative albeit weak relationship with all level-1 variables, except for student valuing of math ($r$ = -.02 to -.07).

At level-2, unweighted bivariate relationships were estimated for 19 predictor variables. The correlation matrix for these variables can be found in Appendix M. Unlike level-1, correlation coefficients of level-2 variables had a wider range, from -.37 between instructional practice-related activities and opportunity to learn number to .70 between ready to teach algebra and ready to teach geometry. As expected, correlation coefficients among the variables measuring the same construct tended to be stronger than those measuring different construct. For example, correlation coefficients ranged from .42 to .70 for ready to teach variables and from .28 to .63 for opportunity to learn variables. Interestingly, it was observed that of the 19 predictors, 10 had negative albeit weak correlation with school resources.

### *Evaluation of HLM Assumptions*

In order to ensure tenability of results produced by multilevel models in this study, an evaluation of HLM assumptions through visual analysis of both level-1 and level-2 random effects of Model 23 was performed. Model 23 was selected because the results of HLM analysis suggested that it was the most efficient model to predict math achievement in South Africa (see HLM Analysis for South Africa).

The data from Figure 45 suggested that level-1 residuals approximated a normal distribution.

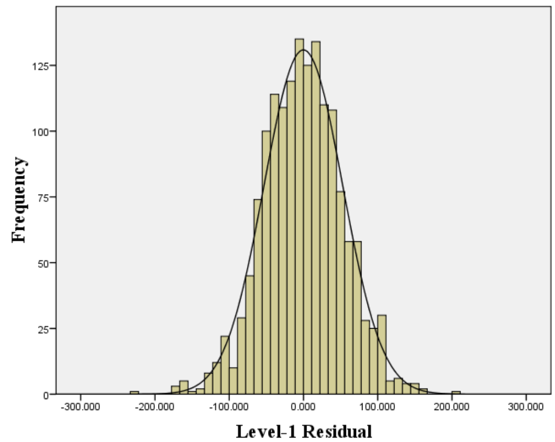



*Figure 45.* Histogram for Level-1 Residuals for South Africa

The scatter plot between level-1 residuals and predicted math achievement, as illustrated in Figure 46, suggested that level-1 variance was not essentially homogeneous. In fact, it appears that there were two clusters of students with one consisting of the majority of students whose predicted math scores were about equal or less than 500 and the other consisting of a small number of students whose predicted math scores were above 500. Interestingly, for the latter group, level-1 residuals were large in magnitude and negative in direction, suggesting their math scores were over predicted.

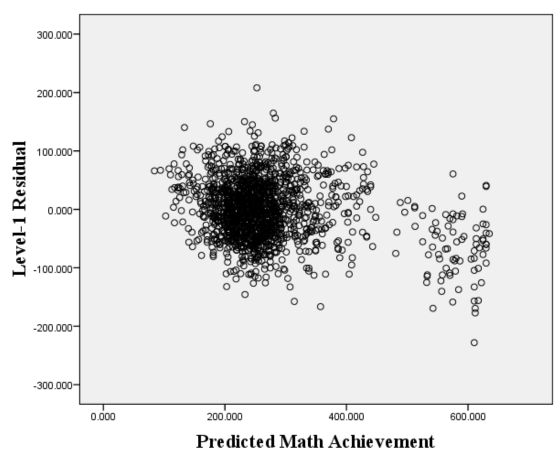



*Figure 46.* Level-1 Residuals by Predicted Math Achievement for South Africa

In order to gain a better understanding of these students, descriptive statistics on level-1 predictors were computed for the two groups of students. The data suggested that there were 55 students whose predicted scores in math were larger than 499. In comparison to their peers whose predicted math scores were less than 499, these students had a higher level of self-confidence in learning math (*M* = 1.65 vs. 1.22), took extra math lessons much less frequently (*M* = .25 vs. 1.58), and had more resources at home for learning math (*M* = 2.69 vs. 1.59).

In order to evaluate the influence of this group of students with over predicted math scores on the results, a decision was made to conduct HLM analysis with two samples: the original sample where all of the students were included and the reduced sample where 55 students with predicted scores larger than 499 were excluded.

The following table compares HLM results of the original sample (*N* = 1,564) with those of the reduced sample (*N* = 1,509) for South Africa. Overall, the results produced by the two datasets were similar. School resources was the only level-2 variable that had significant relationship with math achievement in the final model. It is worth noting that Models 10-23 for the reduced sample were simpler because they were intercept only models (i.e., all the slope variances that were not significant in earlier models were set to 0 in Models 10-23), whereas for the original sample, two slope variances for extra math lessons and self-confidence in learning math were allowed to be random in Models 10-23. However, for the original sample, in the final model, none of these slope variances was significant, which is the same as the final model for the reduced dataset.

In terms of level-1 residuals, it is interesting to see that there remained some students whose math scores were over predicted even though the magnitude of the residuals appeared smaller in the reduced sample than in the original sample. Based on this analysis, the HLM results for the original samples were presented in this study.

Table 78.

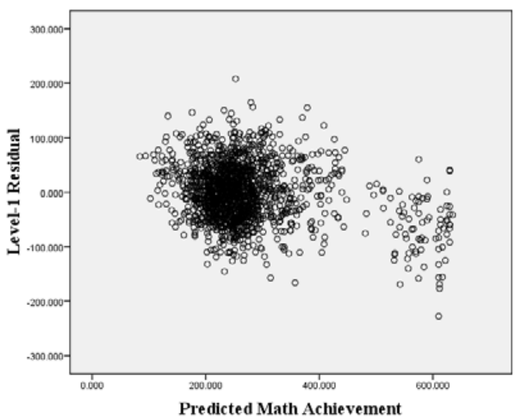
*Comparisons of HLM Results Produced by the Original Sample and Reduced Sample for South Africa*

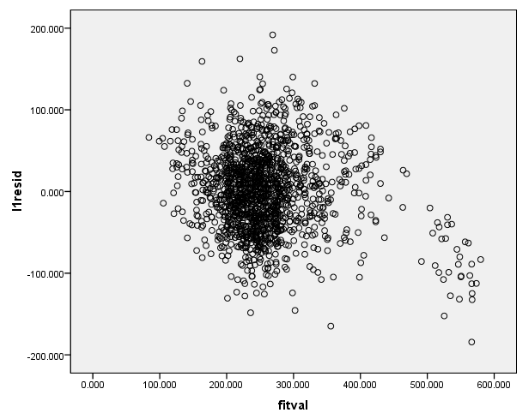
| Model | Original Sample (N = 1564, Highest math score = 670.84) | Reduced Sample (N = 1509, Highest math scores = 499) |
|---|---|---|
| 1 | ICC = .76 | ICC = .69 |
| 2 | Gender: Both fixed and random effects not significant | same |
| 3 | Self-confidence: significant fixed effect, not random effect | same |
| 4 | Valuing math: significant fixed effect, not random effect | Same |
| 5 | Time on HW: Both fixed and random effects not significant | Significant fixed effect, not random |
| 6 | Extra math lessons: Significant fixed and random effects | Significant fixed effect, not random |
| 7 | All student background variables:<br>Gender fixed and random effects not significant<br>Valuing math and time on HW random effects not significant | Same |
| 8 | Home resources: Significant fixed effect, not random effect | Same |
| 9 | All student-related variables: All fixed and random effects significant | All fixed effects significant<br>All slope variances not significant |
| 10 | Opportunity to learn: random coefficient model<br>One cross-level interaction effect significant | Random intercept only model:<br>No interaction effects because all slope variances were set to 0. Level-2 effect not significant |
| 11 | Homework assignment: Not significant | Same |
| 12 | Activities in math lessons: Not significant | Same |
| 13 | Instructional hours: Not significant | Same |
| 14 | Instructional practice: Level-2 not significant Interactions not significant | Same |
| 15 | Preparation to teach: Level-2 not significant Interactions not significant | Same |
| 16 | Ready to teach: Level-2 not significant Interactions not significant | Same |
| 17 | Professional development: Level-2 not significant Interactions not significant | Same |
| 18 | Teacher background model: Level-2 not | Fixed effect for ready to teach number |

Table 78.

*Comparisons of HLM Results Produced by the Original Sample and Reduced Sample for South Africa*

| Model | Original Sample (N = 1564, Highest math score = 670.84) | Reduced Sample (N = 1509, Highest math scores = 499) |
|---|---|---|
| | significant. Interactions not significant | significant |
| 19 | Class size: Level-2 not significant. Interactions not significant | Same |
| 20 | Instructional limitation: Level-2 not significant. Interactions not significant | Same |
| 21 | School resources: Fixed effect for school resources significant | Same |
| 22 | School-background model: Fixed effect for school resources significant | Same |
| 23 | Final model. School resources as the only level-2 variable. Fixed effect for school resources significant | School resources and ready to teach as level-2 variables. Only fixed effect for school resources significant |
| 23 | Level-1 residuals by predicted math scores | Level-1 residuals by predicted math scores |




For level-2 random effects, the empirical Bayes residuals for the intercept and slope as well as empirical Bayes predicted math scores were used to construct the graphs in Figures 47-50. It is worth noting that the model that was used to produce these residuals data consisted of only one level-2 predictor, school resources. Because this variable had three possible values, there were three predicted values in Figures 47 and 50.

As can be seen from Figures 47-48, level-2 intercept residuals appeared to have an approximately normal distribution and homogeneous variance.

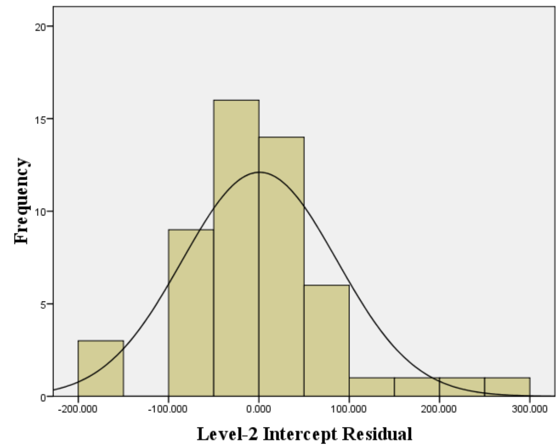



*Figure 47*. Histogram for Level-2 Intercept Residuals for South Africa

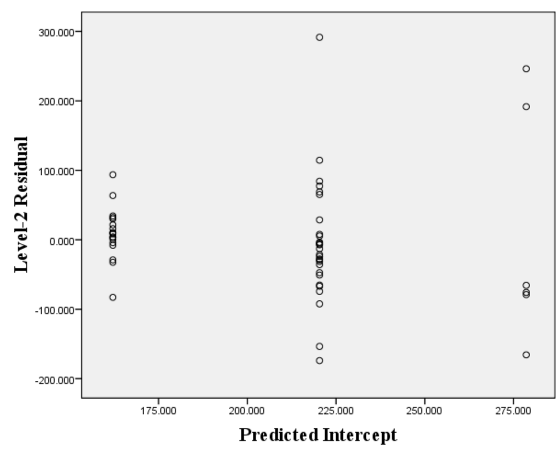



*Figure 48.* Level-2 Intercept Residuals by Predicted Intercept for South Africa

Likewise, as can be seen from Figures 49-50, the slope residuals for extra math lessons also seemed to have an approximately normal distribution and homogeneous variance.

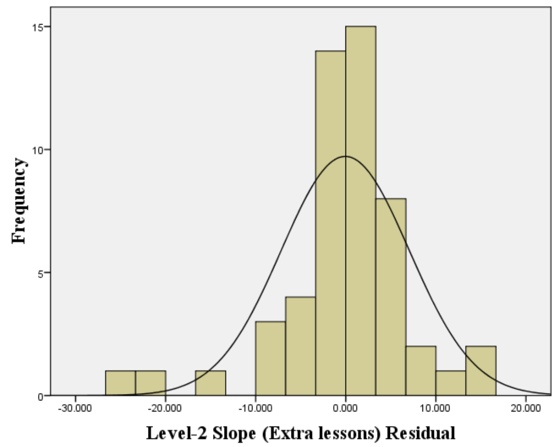



*Figure 49.* Histogram for Level-2 Slope (Extra Lessons) Residuals for South Africa

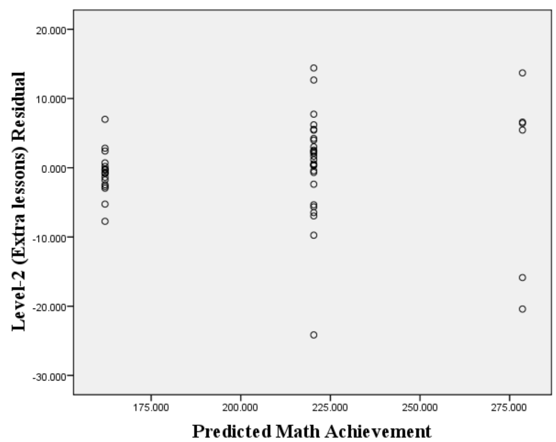



*Figure 50.* Level-2 Slope (Extra Lessons) Residuals by Predicted Math Achievement for South Africa

## *HLM Analysis*

### *Unconditional model (Model 1)*

The HLM analysis started with the unconditional model where none of the level-1 or level-2 predictor was included in the model. The results of the unconditional model are presented in Table 78. For South Africa, the fixed effect for the intercept was 267.57 (*SE* = 17.18, $p$ <.001). The amount of variability in math achievement was significantly different across schools in South Africa ( $\tau_{00}$ = 9,252.67, *SE* = 96.19, $p$ <.001). Within schools, the amount of unexplained variance was much smaller than that between schools

($\sigma^2$ = 2,902.45, *SE* = 53.87). The computed intra-class correlation (ICC) of .76 indicated a relatively strong level of natural clustering of students occurred between schools in South Africa. In other words, approximately 76% of the total variance in math scores occurred between schools.

Table 79.

*Parameter Estimates for Unconditional Model for South Africa*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 1 | | ICC | .76 | | |
| | Fixed | INT | 267.57 | 17.18 | <.001 |
| | Random | $\tau_{00}$ | 9252.67 | 96.19 | <.001 |
| | | $\sigma^2$ | 2902.45 | 53.87 | |

*Research Question 1*

To what extent are student background variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, and tutoring in math) associated with TIMSS 2003 eighth-grade math scores in each country?

In order to answer this research question, first, each of the student background variables was entered separately into Model 1 to predict math achievement. Then, as a group of variables, those that contributed significantly in Models 2-6 were included in Model 7 to predict math achievement. Finally, in order to evaluate model fit in terms proportion of variance accounted for, pseudo $R^2$ was computed for the current model against previously constructed models. Results of these models (Models 2-6) are presented in Table 79.

The data from Table 79 suggested that three out of five fixed effects estimated by Models 2-6 were statistically significant: student self-confidence in learning math ($\gamma$ = 27.53, *SE* = 3.04, *p* <.001), student valuing of math ($\gamma$ = 22.48, *SE* = 3.10, *p* <.001), and extra math lessons in($\gamma$ = -12.81, *SE* = 2.24, *p* <.001). These data suggested that whereas self-confidence in learning math (Model 3) and valuing of math (Model 4) had positive relationships with math achievement, extra math lessons (Model 6) had an inverse relationship with math achievement. Also, in this model, only student self-confidence in learning math and extra math lessons had statistically significant slope variance ($\tau$ = 117.99 and 103.15, *SE* = 10.86 and 10.16, *p* <.001 and .011, respectively). This suggests that schools in South Africa tended to differ significantly in the relationship between math achievement and student self-confidence in learning math and extra math lessons.

An examination of the pseudo $R^2$ values across the five models (Models 2-6) suggested that the addition of individual predictors separately to the unconditional model (Model 1) to predict math achievement resulted in a reduction of the between school variance in three models (i.e., 3%, 9% and 19% in Models 2, 3, and 4, respectively) and an increase of the between school variance in two models (12% and 8% in Models 5 and 6, respectively). For the within school variance, however, some reduction was noted in all the models, between 1% and 13% (see Table 79).

Table 80.
*Parameter Estimates for Models 2-6 (Level-1 Student Background) for South Africa*

| Model | Type | Parameters | Estimates | *SE* | *p* | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|
| 2 | Fixed | INT | 265.72 | 17.00 | <.001 | | |
| | | Gender | 3.55 | 3.44 | .307 | | |
| | Random | $\tau_{00}$ | 9019.34 | 94.97 | <.001 | | |
| | | Gender | 66.14 | 8.13 | .216 | | |
| | | $\sigma^2$ | 2885.06 | 53.71 | | | |
| | Pseudo $R^2$ | | | | | 0.03 | 0.01 |
| 3 | Fixed | INT | 232.96 | 17.39 | <.001 | | |
| | | Self-confidence | 27.53 | 3.04 | <.001 | | |
| | Random | $\tau_{00}$ | 8445.69 | 91.90 | <.001 | | |
| | | Self-confidence | 117.99 | 10.86 | .019 | | |
| | | $\sigma^2$ | 2530.78 | 50.31 | | | |
| | Pseudo $R^2$ | | | | | 0.09 | 0.13 |
| 4 | Fixed | INT | 227.84 | 16.36 | <.001 | | |
| | | Valuing math | 22.48 | 3.10 | <.001 | | |

Table 80.
*Parameter Estimates for Models 2-6 (Level-1 Student Background) for South Africa*

| Model | Type | Parameters | Estimates | *SE* | *p* | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|
| | Random | $\tau_{00}$ | 7486.31 | 86.52 | <.001 | | |
| | | Valuing math | 40.64 | 6.38 | >.500 | | |
| | | $\sigma^2$ | 2774.86 | 52.68 | | | |
| | Pseudo $R^2$ | | | | | 0.19 | 0.04 |
| 5 | Fixed | | 263.65 | 18.26 | <.001 | | |
| | | Homework time | 3.75 | 2.46 | .133 | | |
| | Random | $\tau_{00}$ | 10375.53 | 101.86 | <.001 | | |
| | | Homework time | 39.83 | 6.31 | .355 | | |
| | | $\sigma^2$ | 2887.05 | 53.73 | | | |
| | Pseudo $R^2$ | | | | | -0.12 | 0.01 |
| 6 | Fixed | INT | 283.36 | 17.81 | <.001 | | |
| | | Extra lessons | -12.81 | 2.24 | <.001 | | |
| | Random | $\tau_{00}$ | 9979.98 | 99.90 | <.001 | | |
| | | Extra lessons | 103.15 | 10.16 | .011 | | |
| | | $\sigma^2$ | 2703.82 | 52.00 | | | |
| | Pseudo $R^2$ | | | | | -0.08 | 0.07 |

As a next step of model building, all of the student background variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, and tutoring in math) were included in the combined model, Model 7, to predict math achievement. Notably, in the presence of other level-1 variables in the model, all but gender had statistically significant fixed effects. With fixed effect of -11.97 (*SE* = 1.79, *p* <.001) for extra math lesson, it could be inferred that for each unit increase in extra math lesson (i.e., from 0 for never to 3 for daily), students were expected to reduce 11.97 points in their math scores while controlling for other predictors in the model. Similarly, with fixed effect of 24.67 (*SE* = 2.58, *p* <.001) for self-confidence in learning math, it could be interpreted that for each unit increase in level of self-confidence in learning math (i.e., from 0 for low to 2 for high), it was likely that students would improve 24.67 points in their math scores while controlling for other predictors in the model. Likewise, each unit change in student valuing of math was associated with an increase of 16.24

points (*SE* = 2.80, *p* <.001) in math achievement and each unit change in time students spent on homework was associated with an increase of 4.65 points (*SE* = 2.28, *p* = .046), an after adjusting for other predictors in the model.

Interestingly, it was noted that of all the random effects in this model, only those of extra math lessons and student self-confidence in learning math were statistically significant ($\tau$ = 66.25 and 76.44, *SE* = 8.14 and 8.74, *p* = .010 and .029, respectively). This indicates that the observed relationships between math achievement and extra math lessons and student self-confidence in learning math varied significantly across schools in South Africa.

Table 81.
*Parameter Estimates for Model 7 (Level-1 Student Background) for South Africa*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 7 | Fixed | INT | 217.00 | 18.96 | <.001 |
| | | Gender | 3.58 | 2.86 | .216 |
| | | Extra lessons | -11.97 | 1.79 | <.001 |
| | | Self-confidence | 24.67 | 2.58 | <.001 |
| | | Valuing math | 16.24 | 2.80 | <.001 |
| | | Homework time | 4.65 | 2.28 | .046 |
| | Random | $\tau_{00}$ | 10418.03 | 102.07 | <.001 |
| | | Gender | 41.14 | 6.41 | .205 |
| | | Extra lessons | 66.25 | 8.14 | .010 |
| | | Self-confidence | 76.44 | 8.74 | .029 |
| | | Valuing math | 26.42 | 5.14 | .330 |
| | | Homework time | 26.92 | 5.19 | >.500 |
| | | $\sigma^2$ | 2286.76 | 47.82 | |

An evaluation of model fit was also conducted between Model 7 and previously constructed models, Models 2-6. As can be seen from Table 81, the inclusion of student background variables in Model 7 resulted in some reduction in the amount of variance accounted for in math achievement within schools (from 10% when compared with Model 3 to 21% when compared with Models 2 and 5). Between schools, the amount of variance appeared to increase notably in Models 2 to 4, from 16% 39%. In sum, Model 7 was more efficient than earlier models in that it accounted for more variance in math achievement within schools in South Africa. However, this model appeared to be less efficient than previously constructed models in that it accounted for less variance in math achievement between schools in South Africa.

Table 82.
*Comparison of $R^2$ between Model 7 and Previously Constructed Models for South Africa*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 7 vs. 2 | -0.16 | 0.21 |
| 7 vs. 3 | -0.23 | 0.10 |
| 7 vs. 4 | -0.39 | 0.18 |
| 7 vs. 5 | 0.00 | 0.21 |
| 7 vs. 6 | -0.04 | 0.15 |

### *Research Question 2*

To what extent are home resources variables (i.e., availability of calculator, computer, and desk for student use) associated with TIMSS 2003 eighth-grade math scores in each country?

When the level-1 predictor home resources was added to the unconditional model to predict math achievement, a reduction of 1% in the within school variance and 25% in the between school variance was noted (see Table 82). As a fixed effect, home resources was statistically significant ($\gamma$ = 7.55, *SE* = 2.30, *p* = .002), meaning that each unit change in home resources (i.e., from 0 to 3) was associated with an increase of 7.55 points in student math achievement, while not controlling for other variables in the model. As a random effect, home resources was not statistically significant ($\tau$ = 53.80, *SE* = 7.33, *p* = .118). This indicates that the relationship between home resources and math achievement did not vary significantly across schools in South Africa.

Table 83.
*Parameter Estimates for Level-1 Home Resources Model for South Africa*

| Model | Type | Parameters | Estimates | *SE* | *p* | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|
| 8 | Fixed | INT | 253.63 | 15.49 | <.001 | | |
| | | Home resources | 7.55 | 2.30 | .002 | | |
| | Random | $\tau_{00}$ | 6957.50 | 83.41 | <.001 | | |
| | | Home resources | 53.80 | 7.33 | .118 | | |
| | | $\sigma^2$ | 2860.28 | 53.48 | | | |
| | Pseudo $R^2$ | | | | | 0.25 | 0.01 |

Given the findings obtained from Models 7 and 8, five out of six student-related variables were entered into the unconditional model to make Model 9. Gender was excluded from Model 9 because both of its fixed and random effects were not statistically significant in Model 7. Also, in Model 9, all the slope variances for student valuing of math and time students spent on homework were set to 0 because they were not statistically significant in earlier models.

As can be seen from Table 83, all of the level-1 variables had statistically significant fixed effects. Of the five variables in the models, student self-confidence in learning math appears to have the strongest positive relationship with math achievement ($\gamma$ = 24.17, *SE* = 2.72, *p* <.001). Following was student valuing of math with a fixed effect of 16.23 (*SE* = 2.86, *p* <.001). Next were home resources ($\gamma$ = 6.81, *SE* = 1.92, *p* <.001) and time student spent on homework ($\gamma$ = 5.33, *SE* = 2.22, *p* = .017). As for extra math lesson, an inverse relationship was noted between this predictor and math achievement ($\gamma$ = -12.54, *SE* = 1.81, *p* <.001). These results suggest that the more self-confidence students expressed in learning math, the higher students valued math, the more time students spent on their homework, and the more home resources students had at home was associated with the higher math scores students tended to achieve. In contrast, it appears that the more frequently students took extra math lessons, the poorer math scores they seem to earn.

In terms of random effects, all were found statistically significant, suggesting that a considerable amount of variance between schools remained to be explained. In addition, it could be inferred that the relationship between math achievement and extra math lessons and student self-confidence in learning math varied significantly across schools.

As compared to Models 7 and 8, Model 9 appeared more efficient because it accounted for a significantly higher amount of the explained variance between schools (i.e., 7% when compared to Model 7). In terms of the explained variance within schools, an increase of up to 20% was observed when compared with Model 8. As a result of these comparisons, Model 9 was selected as the foundational level-1 model for further examination of the relationships between level-2 predictors and math achievement.

Table 84.
*Parameter Estimates for Combined Level-1 Predictors Model for South Africa*

| Model | Type | Parameters | Estimates | SE | p | Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|
| 9 | Fixed | INT | 208.39 | 19.23 | <.001 | | | |
| | | Extra lessons | -12.54 | 1.81 | <.001 | | | |
| | | Self-confidence | 24.17 | 2.72 | <.001 | | | |
| | | Valuing math | 16.23 | 2.86 | <.001 | | | |
| | | Homework time | 5.33 | 2.22 | .017 | | | |
| | | Home resources | 6.81 | 1.92 | .001 | | | |
| | Random | $\tau_{00}$ | 9644.91 | 98.21 | <.001 | | | |
| | | Extra lessons | 68.25 | 8.26 | .045 | | | |
| | | Self-confidence | 69.04 | 8.31 | .034 | | | |
| | | $\sigma^2$ | 2292.92 | 47.88 | | | | |
| | Pseudo $R^2$ | | | | | 7 | 0.07 | 0.00 |
| | | | | | | 8 | -0.39 | 0.20 |

*Research Question 3*

To what extent are instructional variables (i.e., opportunity to learn, activities in math lessons, amount of homework assignment, and instructional time) associated with TIMSS 2003 eighth-grade math scores in each country?

In addressing this research question, a similar strategy for model building used in Research Question 1 was applied here. That is, each of the level-2 instructional practice variables was first added to the foundational level-1 model (Model 9) to make Models 10-13. Then, as a group, those variables with significant fixed effects in Models 10-13 were included in the combined instructional practices model, Model 14. It is important to note that in these models, all possible cross-level interactions between level-1 and level-2 predictors were allowed. The results of Models 10-14 are presented in Tables 84-87.

As can be seen in Table 84, Model 10 produced one statistically significant cross-level interaction between opportunity to learn data and student self-confidence in learning math ($\gamma$ = -.14, SE = .05, $p < .015$). Also, in this model, all of the fixed effects, except for those of level-2 predictors, were statistically significant. This means that, whereas all of the level-1 predictors appear to be good predictors of math achievement in South Africa, there was not enough evidence to make a similar statement about the opportunity to learn variables.

Similar to Model 10, none of the level-2 predictors was found statistically significant in Models 11-13. In contrast, all of the level-1 predictors, except for extra math lessons and student self-confidence in learning math in Model 12, were statistically significant. In terms of random effects, the slope variance of student self-confidence in learning math in Model 11 was the only one that was not statistically significant ($\tau$ =

63.16, *SE* = 7.95, *p* = .050). This means, in South Africa, the relationship between math achievement and self-confidence did not appear to differ from one school to another.

Table 85.
*Parameter Estimates for Level-2 Instructional Practices Models for South Africa*

| Model | Type | Parameters | Estimates | SE | p |
|---|---|---|---|---|---|
| 10 | Fixed | INT | 212.72 | 55.63 | .001 |
| | | Opportunity_algebra | -0.15 | 0.57 | .798 |
| | | Opportunity_data | 0.19 | 0.65 | .775 |
| | | Opportunity_geometry | -1.42 | 0.72 | .055 |
| | | Opportunity_measurement | 1.12 | 0.83 | .183 |
| | | Opportunity_number | 0.08 | 0.74 | .918 |
| | | Extra lessons | -18.20 | 6.28 | .006 |
| | | Opportunity_algebra * Extra lessons | -0.02 | 0.07 | .753 |
| | | Opportunity_data * Extra lessons | 0.03 | 0.06 | .647 |
| | | Opportunity_geometry * Extra lessons | 0.11 | 0.08 | .172 |
| | | Opportunity_measurement * Extra lessons | -0.13 | 0.08 | .102 |
| | | Opportunity_number * Extra lessons | 0.10 | 0.09 | .267 |
| | | Self-confidence | 28.12 | 8.20 | .002 |
| | | Opportunity_algebra * Self-confidence | 0.09 | 0.09 | .334 |
| | | Opportunity_data * Self-confidence | -0.14 | 0.05 | .015 |
| | | Opportunity_geometry * Self-confidence | 0.10 | 0.09 | .268 |
| | | Opportunity_measurement * Self-confidence | -0.11 | 0.09 | .221 |
| | | Opportunity_number * Self-confidence | -0.04 | 0.12 | .766 |
| | | Valuing math | 15.55 | 3.00 | <.001 |
| | | Homework time | 5.45 | 2.19 | .013 |
| | | | 6.78 | 1.92 | .001 |
| | Random | $\tau_{00}$ | 7986.85 | 89.37 | <.001 |
| | | Extra lessons | 60.36 | 7.77 | .020 |
| | | Self-confidence | 26.38 | 5.14 | .205 |
| | | $\sigma^2$ | 2299.31 | 47.95 | |
| 11 | Fixed | INT | 181.83 | 23.53 | <.001 |
| | | Homework assignment | 25.05 | 16.76 | .141 |
| | | Extra lessons | -10.14 | 2.74 | .001 |
| | | Homework assignment * Extra lessons | -2.24 | 2.24 | .323 |
| | | Self-confidence | 19.10 | 3.92 | <.001 |
| | | Homework assignment * Self-confidence | 4.70 | 2.46 | .062 |
| | | Valuing math | 16.01 | 2.86 | <.001 |
| | | Homework time | 5.43 | 2.21 | .014 |
| | | Home resources | 6.70 | 1.93 | .001 |
| | Random | $\tau_{00}$ | 9574.97 | 97.85 | <.001 |
| | | Extra lessons | 68.81 | 8.29 | .038 |
| | | Self-confidence | 63.16 | 7.95 | .050 |
| | | $\sigma^2$ | 2294.04 | 47.90 | |
| 12 | Fixed | INT | 434.97 | 210.43 | .044 |
| | | Content_activities | 27.10 | 81.45 | .741 |

Table 85.
*Parameter Estimates for Level-2 Instructional Practices Models for South Africa*

| Model | Type | Parameters | Estimates | SE | p |
|---|---|---|---|---|---|
| | | Instruction_activities | -132.11 | 100.59 | .195 |
| | | Extra lessons | -25.45 | 21.99 | .253 |
| | | Content_activities * Extra lessons | 9.79 | 7.15 | .177 |
| | | Instruction_activities * Extra lessons | -2.57 | 9.65 | .791 |
| | | Self-confidence | 15.11 | 24.65 | .542 |
| | | Content_activities * Self-confidence | -8.73 | 10.37 | .404 |
| | | Instruction_activities * Self-confidence | 12.17 | 14.78 | .414 |
| | | Valuing math | 15.73 | 2.99 | <.001 |
| | | Homework time | 4.91 | 2.27 | .031 |
| | | Home resources | 6.77 | 1.89 | .001 |
| | Random | $\tau_{00}$ | 9238.02 | 96.11 | <.001 |
| | | Extra lessons | 82.26 | 9.07 | .032 |
| | | Self-confidence | 68.49 | 8.28 | .038 |
| | | $\sigma^2$ | 2286.29 | 47.82 | |
| 13 | Fixed | INT | 208.97 | 19.30 | <.001 |
| | | Instructional hours | -0.14 | 0.18 | .460 |
| | | Extra lessons | -12.62 | 1.82 | <.001 |
| | | Instructional hours * Extra lessons | 0.00 | 0.03 | .903 |
| | | Self-confidence | 24.18 | 2.72 | <.001 |
| | | Instructional hours * Self-confidence | -0.01 | 0.03 | .862 |
| | | Valuing math | 16.26 | 2.85 | <.001 |
| | | Homework time | 5.28 | 2.21 | .017 |
| | | Home resources | 6.79 | 1.93 | .001 |
| | Random | $\tau_{00}$ | 9738.81 | 98.69 | <.001 |
| | | Extra lessons | 68.68 | 8.29 | .038 |
| | | Self-confidence | 72.85 | 8.54 | .028 |
| | | $\sigma^2$ | 2293.84 | 47.89 | |

In terms of model fit, in comparison with the foundational level-1 model (Model 9), Model 10 appeared to be the most efficient model because the amount of explained variance between schools in this model increased by 17% (see Table 85). As for the within school variance, no significant difference was observed between Models 10-13 and Model 9 (pseudo $R^2 = 0$).

Table 86.
*Comparison of $R^2$ between Level-2 Instructional Practice Models and Foundational Level-1 Model*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 10 vs. 9 | 0.17 | 0.00 |
| 11 vs. 9 | 0.01 | 0.00 |
| 12 vs. 9 | 0.04 | 0.00 |
| 13 vs. 9 | 0.07 | 0.00 |

Figure 51 presents the nature of the interaction between student self-confidence in learning math and opportunity to learn data produced by Model 10. For students who expressed a low or medium self-confidence in learning math, increased opportunity to learn data appeared to be associated with increased student math scores. In contrast, for students with a high level of self-confidence in learning math, increased opportunity to learn data seems to be associated with lower math scores for the students. Thus, it looks like that high opportunity to learn data worked best for students with a low level of self-confidence in learning math. However, in comparing the three groups of students, it appears that those reported to have a high level of self-confidence in learning math consistently outperformed their peers who reported having a low or medium level of self-confidence in learning math. The sizes of differences in math achievement among these three groups of students, however, were large when there was little opportunity to learn data and became significantly smaller when there was higher opportunity to learn data.

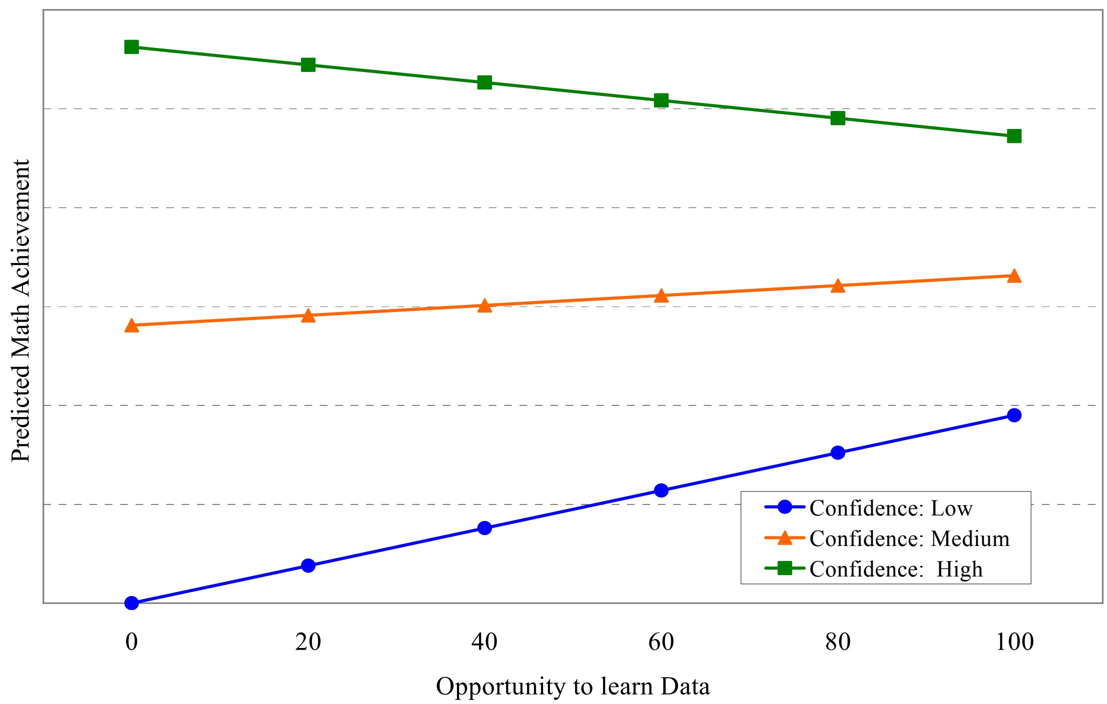



*Figure 51.* Interaction between Opportunity to Learn Data and Student Self-confidence in Learning Math in South Africa

Similar to what was observed in Models 10-13, when using all the level-2 instructional practice variables to predict math achievement, Model 14 did not produce any statistically significant cross-level interaction effect as well as level-2 fixed effects (see Table 86). This means that, in South Africa, instructional practices-related predictors did not appear to contribute significantly in the statistical prediction of student math achievement. As for level-1 fixed effects, three were found statistically significant: valuing of math ($\gamma$ = 15.03, *SE* = 3.08, *p* <.001), time students spent on homework ($\gamma$ = 5.08, *SE* = 2.22, *p* = .022), and home resources ($\gamma$ = 6.73, *SE* = 1.91, *p* = .001). These results suggested that increased home resources, or time students spent on homework, or valuing of math tended to be associated with increased student math scores. As for random effects, whereas the variances of the intercept ($\tau_{00}$ = 8146.44, *SE* = 90.26, *p* <.001) and the slope of extra math lessons (τ = 67.99, *SE* = 8.25, *p* = .016) were statistically significant, that of the slope of student self-confidence in learning math was not significant (τ = 31.96, *SE* = 5.65, *p* = .152). This suggested that schools in South Africa differed significantly in math achievement as well as the relationship between math achievement and extra math lessons.

Table 87.
*Parameter Estimates for the Combined Level-2 Instructional Practices Model for South Africa*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 14 | Fixed | INT | 350.31 | 177.78 | .055 |
| | | Homework assignment | 28.07 | 20.06 | .169 |
| | | Opportunity_algebra | -0.24 | 0.60 | .689 |
| | | Opportunity_data | 0.23 | 0.62 | .717 |
| | | Opportunity_geometry | -1.31 | 0.68 | .062 |
| | | Opportunity_measurement | 1.09 | 0.74 | .151 |
| | | Opportunity_number | -0.31 | 0.72 | .667 |
| | | Content_activities | -8.22 | 85.80 | .925 |
| | | Instruction_activities | -57.18 | 80.46 | .481 |
| | | Instructional hours | -0.08 | 0.23 | .732 |
| | | Extra lessons | -21.92 | 19.05 | .257 |
| | | Homework assignment * Extra lessons | -3.72 | 2.40 | .128 |
| | | Opportunity_algebra * Extra lessons | 0.02 | 0.07 | .783 |
| | | Opportunity_data * Extra lessons | 0.02 | 0.05 | .768 |
| | | Opportunity_geometry * Extra lessons | 0.11 | 0.08 | .172 |
| | | Opportunity_measurement * Extra lessons | -0.15 | 0.08 | .063 |
| | | Opportunity_number * Extra lessons | 0.09 | 0.08 | .290 |
| | | Content_activities * Extra lessons | 13.62 | 7.71 | .084 |
| | | Instruction_activities * Extra lessons | -8.82 | 9.34 | .351 |
| | | Instructional hours * Extra lessons | -0.02 | 0.03 | .535 |
| | | Self-confidence | 24.95 | 25.50 | .334 |
| | | Homework assignment * Self-confidence | 4.60 | 2.95 | .126 |
| | | Opportunity_algebra * Self-confidence | 0.07 | 0.10 | .468 |
| | | Opportunity_data * Self-confidence | -0.12 | 0.06 | .050 |
| | | Opportunity_geometry * Self-confidence | 0.09 | 0.09 | .315 |
| | | Opportunity_measurement * Self-confidence | -0.10 | 0.09 | .299 |
| | | Opportunity_number * Self-confidence | -0.07 | 0.11 | .494 |
| | | Content_activities * Self-confidence | -2.17 | 7.12 | .762 |
| | | Instruction_activities * Self-confidence | 2.47 | 11.06 | .824 |
| | | Instructional hours * Self-confidence | 0.00 | 0.03 | .966 |
| | | Valuing math | 15.03 | 3.08 | <.001 |
| | | Homework time | 5.08 | 2.22 | .022 |
| | | Home resources | 6.73 | 1.91 | .001 |
| | Random | $\tau_{00}$ | 8146.44 | 90.26 | <.001 |
| | | Extra lessons | 67.99 | 8.25 | .016 |

Table 87.
*Parameter Estimates for the Combined Level-2 Instructional Practices Model for South Africa*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Self-confidence | 31.96 | 5.65 | .152 |
| | | $\sigma^2$ | 2296.95 | 47.93 | |

As evident in Table 87, the amount of explained variance between schools in Model 14 appeared more significant than all of the compared models, except for Model 10. Specifically, when using Model 14 instead of Models 9 or 13, an increase of 16% in the explained variance between schools was likely to result. However, changing from Models 10 to Model 14 seemed to reduce the amount of explained variance between schools by 2%. In terms of the variance within schools, no change in the variance was noted across these models. Thus, in consideration of the amount of the explained variance between schools, Model 10 serves as the most efficient instructional practice model to predict math achievement.

Table 88.
*Comparison of* $R^2$ *between Model 14 and Previously Constructed Models 9-13 for South Africa*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 9 | 0.16 | 0.00 |
| 10 | -0.02 | 0.00 |
| 11 | 0.15 | 0.00 |
| 12 | 0.12 | 0.00 |
| 13 | 0.16 | 0.00 |

### *Research Question 4*

To what extent are teacher-related variables (i.e., preparation to teach, ready to teach, and professional development) associated with TIMSS 2003 eighth-grade math scores in each country?

Similarly, incremental model building strategies were applied to examine the relationships among teacher-related variables (i.e., preparation to teach, ready to teach, and professional development) and math achievement. Results of these models (Models 15-18) are presented in Table 88-90.

The data in Table 88 show that none of the teacher background-related variables was statistically significant. However, for level-1 fixed effects, all were statistically significant, except for that of student self-confidence in learning math in Model 16 ($\gamma$ = 18.82, *SE* = 11.81, *p* = .118). These data suggest that student-related predictors tended to have stronger relationships with math achievement than teacher-related predictors across South African schools. Similarly, it was noted that all of the random effects were statistically significant in these models, except for that of student self-confidence in learning math in Model 15 ($\tau$ = 56.04, *SE* = 7.49, *p* = .068). It could be inferred from these results that a significant amount of variance in math achievement remained unexplained, both within and between schools.

When comparing the proportion of variance accounted for by Models 15-17 with that of the foundational level-1 model (Model 9), it appears that Model 15 was the most efficient one (see Table 88). Specifically, whereas the inclusion of readiness to teach math content in Model 16 and math-related professional development in Model 17 resulted in a reduction of 1-4% in the amount of explained variance between schools, the addition of preparation to teach math content in Model 15 resulted in an increase of 4% of the explained between school variance. Further, no improvement in the within school variance was noted by use of these models instead of Model 9.

Table 89.
*Parameter Estimates for Teacher Background Models for South Africa*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 15 | Fixed | INT | 247.45 | 43.71 | <.001 |
| | | Preparation | -50.91 | 47.04 | .285 |
| | | Extra lessons | -14.67 | 2.96 | <.001 |
| | | Preparation * Extra lessons | 2.76 | 3.63 | .450 |

Table 89.
*Parameter Estimates for Teacher Background Models for South Africa*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Self-confidence | 18.57 | 6.67 | .008 |
| | | Preparation * Self-confidence | 7.69 | 6.90 | .271 |
| | | Valuing math | 16.16 | 2.84 | <.001 |
| | | Homework time | 5.42 | 2.22 | .015 |
| | | Home resources | 6.82 | 1.93 | .001 |
| | Random | $\tau_{00}$ | 9272.88 | 96.30 | <.001 |
| | | Extra lessons | 66.09 | 8.13 | .043 |
| | | Self-confidence | 56.04 | 7.49 | .068 |
| | | $\sigma^2$ | 2296.36 | 47.92 | |
| 16 | Fixed | INT | 147.83 | 78.60 | .066 |
| | | Ready_number | -6.77 | 41.09 | .870 |
| | | Ready_algebra | 16.38 | 18.91 | .391 |
| | | Ready_measurement | 33.37 | 27.62 | .233 |
| | | Ready_geometry | 0.43 | 14.52 | .977 |
| | | Ready_data | -4.59 | 33.36 | .892 |
| | | Extra lessons | -20.54 | 8.36 | .018 |
| | | Ready_number * Extra lessons | 5.33 | 4.40 | .232 |
| | | Ready_algebra * Extra lessons | 1.40 | 1.77 | .434 |
| | | Ready_measurement * Extra lessons | -1.73 | 2.87 | .549 |
| | | Ready_geometry * Extra lessons | -2.14 | 1.83 | .249 |
| | | Ready_data * Extra lessons | 1.49 | 2.90 | .610 |
| | | Self-confidence | 18.82 | 11.81 | .118 |
| | | Ready_number * Self-confidence | 0.95 | 7.09 | .895 |
| | | Ready_algebra * Self-confidence | 1.51 | 5.60 | .789 |
| | | Ready_measurement * Self-confidence | -3.41 | 5.01 | .499 |
| | | Ready_geometry * Self-confidence | 0.39 | 4.18 | .927 |
| | | Ready_data * Self-confidence | 4.63 | 5.64 | .416 |
| | | Valuing math | 15.60 | 2.96 | <.001 |
| | | Homework time | 5.14 | 2.17 | .018 |
| | | Home resources | 6.88 | 1.94 | .001 |
| | Random | $\tau_{00}$ | 10042.14 | 100.21 | <.001 |
| | | Extra lessons | 80.78 | 8.99 | .024 |
| | | Self-confidence | 81.36 | 9.02 | .024 |
| | | $\sigma^2$ | 2288.86 | 47.84 | |
| 17 | Fixed | INT | 195.86 | 17.71 | <.001 |
| | | Professional development | 3.95 | 8.55 | .646 |
| | | Extra lessons | -12.42 | 1.90 | <.001 |
| | | Professional development * Extra lessons | -0.02 | 0.92 | .981 |
| | | Self-confidence | 27.31 | 3.84 | <.001 |
| | | Professional development * Self-confidence | -1.01 | 1.33 | .451 |
| | | Valuing math | 16.26 | 2.88 | <.001 |
| | | Homework time | 5.32 | 2.22 | .017 |
| | | Home resources | 6.77 | 1.91 | .001 |
| | Random | $\tau_{00}$ | 9723.46 | 98.61 | <.001 |

Table 89.
*Parameter Estimates for Teacher Background Models for South Africa*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Extra lessons | 69.35 | 8.33 | .037 |
| | | Self-confidence | 67.66 | 8.23 | .038 |
| | | $\sigma^2$ | 2294.58 | 47.90 | |

Similar to the results observed in recent models, Model 18 with all of the teacher-related variables included in the model (i.e., preparation to teach, ready to teach math topics, and math-related professional development) yielded neither statistically significant cross-level interaction effects nor statistically significant level-2 main effects (see Table 89). What were found statistically significant in this model were four level-1 fixed effects and all of the random effects, except for that of student self-confidence in learning math ($\tau$ = 65.08, *SE* = 8.07, *p* = .052).

Of the level-1 significant effects, extra math lessons appears to the have the strongest yet negative relationship with math achievement ($\gamma$ = -22.57, *SE* = 7.41, *p* = .004). Following was student valuing of math with a positive fixed effect of 15.58 (*SE* = 2.97, *p* <.001). Next were home resources and time students spent on homework ($\gamma$ = 6.82 and 5.17, *SE* = 1.93 and 2.18, *p* = .001 and .018). These results suggest that for every unit increases in extra math lessons, student math achievement tended to reduce by 22.57 points. In contrast, with every unit increase in student valuing of math, students tended to improve 15.58 points in their math scores. Similarly, the more home resources student had to support their learning, the better they tended to perform in math. Likewise, the more time students engaged in homework, the higher math scores they tended to achieve. As for the random effects, it could be inferred that a considerable amount of variance remained to be explained within and between schools in South Africa.

Table 90.
*Parameter Estimates for the Combined Teacher Background Model for South Africa*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 18 | Fixed | INT | 163.45 | 78.70 | .043 |
| | | Preparation | -83.39 | 49.15 | .096 |
| | | Professional development | 3.98 | 7.96 | .619 |
| | | Ready_number | 7.27 | 41.40 | .862 |
| | | Ready_algebra | -8.19 | 21.07 | .699 |
| | | Ready_measurement | 33.79 | 18.10 | .068 |
| | | Ready_geometry | 23.03 | 19.16 | .236 |
| | | Ready_data | 5.46 | 23.51 | .817 |
| | | Extra lessons | -22.57 | 7.41 | .004 |
| | | Preparation * Extra lessons | 3.07 | 3.63 | .403 |
| | | Professional development * Extra lessons | 0.23 | 0.92 | .804 |
| | | Ready_number * Extra lessons | 4.97 | 3.81 | .199 |
| | | Ready_algebra * Extra lessons | 3.25 | 2.81 | .254 |
| | | Ready_measurement * Extra lessons | -2.04 | 2.59 | .435 |
| | | Ready_geometry * Extra lessons | -3.37 | 2.05 | .106 |
| | | Ready_data * Extra lessons | 0.81 | 2.52 | .749 |
| | | Self-confidence | 20.58 | 13.48 | .134 |
| | | Preparation * Self-confidence | 8.69 | 7.17 | .232 |
| | | Professional development * Self-confidence | -1.38 | 1.29 | .289 |
| | | Ready_number * Self-confidence | -0.94 | 7.61 | .903 |
| | | Ready_algebra * Self-confidence | 2.30 | 5.60 | .683 |
| | | Ready_measurement * Self-confidence | -2.00 | 4.39 | .651 |
| | | Ready_geometry * Self-confidence | -0.97 | 4.49 | .830 |
| | | Ready_data * Self-confidence | 3.43 | 4.90 | .488 |
| | | Valuing math | 15.58 | 2.97 | <.001 |
| | | Homework time | 5.17 | 2.18 | .018 |
| | | Home resources | 6.82 | 1.93 | .001 |
| | Random | $\tau_{00}$ | 9279.73 | 96.33 | <.001 |
| | | Extra lessons | 84.52 | 9.19 | .019 |
| | | Self-confidence | 65.08 | 8.07 | .052 |
| | | $\sigma^2$ | 2292.84 | 47.88 | |

In terms of model fit, the results in Table 90 show that Model 18 worked equally well as Model 15 but better than Models 9, 16 and 17. Whereas there was no difference in the amount of the variance between schools accounted for by Models 15 and 18, there was an increase of 4% to 8% in the amount of variance accounted for between schools when using Model 18 as opposed to Models 9, 16, or 17.

Table 91.
*Comparison of* $R^2$ *between Model 18 and Previously Constructed Models 9 and 15-17 for South Africa*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 9 | 0.04 | 0.00 |
| 15 | 0.00 | 0.00 |
| 16 | 0.08 | 0.00 |
| 17 | 0.05 | 0.00 |

*Research Question 5*

To what extent are school-related variables (i.e., class size, school resources for math instruction, and teacher perception of math instructional limitation due to student factor) associated with TIMSS 2003 eighth-grade math scores in each country?

Table 91 provides a summary of the results for Models 19-21 where school-related variables (i.e., class size, school resources for math instruction, and teacher perception of math instructional limitations due to student factors) were separately included in the models to predict math achievement. In these models, no statistically significant cross-level interaction effects were detected. However, there was one statistically significant level-2 main effect: school resources for math instruction in Model 21. With the fixed effect of 57.60 ($SE$ = 26.97, $p$ = .037), it could be inferred that for every unit increased in school resources for math instruction, students were expected to improve their math scores by 57.60 points, after adjusting for other predictors in the model. Similarly, with all of the level-1 fixed effects significant, it could be interpreted that student-related variables contributed significantly in the prediction of math achievement in South Africa.

Table 92.
*Parameter Estimates for School Background Models for South Africa*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 19 | Fixed | INT | 237.92 | 72.48 | .002 |
| | | Class size | -12.77 | 27.26 | .641 |
| | | Extra lessons | -13.19 | 6.30 | .041 |

Table 92.
*Parameter Estimates for School Background Models for South Africa*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| | | Class size * Extra lessons | 0.35 | 2.45 | .887 |
| | | Self-confidence | 28.12 | 6.53 | <.001 |
| | | Class size * Self-confidence | -1.59 | 2.92 | .589 |
| | | Valuing math | 16.36 | 2.91 | <.001 |
| | | Homework time | 5.35 | 2.23 | .017 |
| | | Home resources | 6.79 | 1.93 | .001 |
| | Random | $\tau_{00}$ | 9590.79 | 97.93 | <.001 |
| | | Extra lessons | 64.31 | 8.02 | .034 |
| | | Self-confidence | 71.39 | 8.45 | .034 |
| | | $\sigma^2$ | 2293.58 | 47.89 | |
| 20 | Fixed | INT | 208.91 | 27.95 | <.001 |
| | | Instructional limitation | -0.59 | 20.89 | .978 |
| | | Extra lessons | -13.35 | 2.64 | <.001 |
| | | Instructional limitation * Extra lessons | 0.85 | 2.41 | .727 |
| | | Self-confidence | 24.83 | 3.35 | <.001 |
| | | Instructional limitation * Self-confidence | -0.65 | 2.24 | .772 |
| | | Valuing math | 16.21 | 2.87 | <.001 |
| | | Homework time | 5.30 | 2.23 | .018 |
| | | Home resources | 6.83 | 1.95 | .001 |
| | Random | $\tau_{00}$ | 9828.80 | 99.14 | <.001 |
| | | Extra lessons | 69.71 | 8.35 | .035 |
| | | Self-confidence | 71.24 | 8.44 | .029 |
| | | $\sigma^2$ | 2295.12 | 47.91 | |
| 21 | Fixed | INT | 161.37 | 17.91 | <.001 |
| | | School resources | 57.60 | 26.97 | .037 |
| | | Extra lessons | -8.58 | 2.43 | .001 |
| | | School resources * Extra lessons | -4.83 | 2.58 | .066 |
| | | Self-confidence | 23.94 | 2.63 | <.001 |
| | | School resources * Self-confidence | 0.47 | 2.65 | .860 |
| | | Valuing math | 16.34 | 2.88 | <.001 |
| | | Homework time | 5.34 | 2.21 | .016 |
| | | Home resources | 6.74 | 1.88 | .001 |
| | Random | $\tau_{00}$ | 8319.70 | 91.21 | <.001 |
| | | Extra lessons | 57.07 | 7.55 | .019 |
| | | Self-confidence | 71.57 | 8.46 | .028 |
| | | $\sigma^2$ | 2294.88 | 47.90 | |

In comparing Models 19-21 with Model 9 in terms of the proportion of variance accounted for, it looks like that Model 21 worked the best (see Table 92). Specifically, the use of Model 21 as opposed to Model 9 increased the amount of between school variance accounted for by 14%; whereas the use of Model 19 resulted in an increase in the between school variance by 1% compared to Model 9, and the use of Model 20 resulted in a reduction of 2% in the between school variance compared to Model 9.

Table 93.
*Comparison of* $R^2$ *between Level-2 Teacher Background and Foundational Level-1 Model for South Africa*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 19 vs. 9 | 0.01 | 0.00 |
| 20 vs. 9 | -0.02 | 0.00 |
| 21 vs. 9 | 0.14 | 0.00 |

Similar to Model 21, Model 22 with all of the school background-related predictors did not produce any statistically significant cross-level interaction effects (see Table 93). However, as a level-2 predictor, school resources was statistically significant. With the fixed effect of 55.75 (*SE* = 25.34, *p* = .033), it could be inferred that each unit increased in school resources for math instruction, student math scores were expected to increase by 57.60 points, after adjusting for other predictors in the model. Interestingly, in the presence of all the school-related predictors, extra math lessons became statistically insignificant ($\gamma$ = -8.77, *SE* = 5.28, *p* = .103). All other level-1 predictors in the model appeared to have statistically significant contributions to the prediction of math achievement. With the fixed effects of 5.35 (*SE* = 2.22, *p* = .016) for time students spent on homework to 28.00 (*SE* = 6.68, *p* < .01) for student self-confidence in learning math, it could be interpreted that students could increase their math scores by 5.35 to 28 points for each unit increased in the corresponding predictor, while controlling for other predictors in the model.

Again, with all the random effects were found statistically significant in Model 22, it could be conclude that a significant amount of variance within and between schools in South Africa remained to be explained.

Table 94.
*Parameter Estimates for the Combined School Background Model for South Africa*

| Model | Type | Parameters | Estimates | SE | p |
|---|---|---|---|---|---|
| 22 | Fixed | INT | 181.65 | 53.91 | .002 |
| | | Instructional limitation | 0.43 | 19.65 | .983 |
| | | Class size | -8.41 | 21.03 | .690 |
| | | School resources | 55.75 | 25.34 | .033 |
| | | Extra lessons | -8.77 | 5.28 | .103 |
| | | Instructional limitation * Extra lessons | 0.87 | 2.31 | .709 |
| | | Class size * Extra lessons | -0.27 | 2.06 | .896 |
| | | School resources * Extra lessons | -4.58 | 2.46 | .068 |
| | | Self-confidence | 28.00 | 6.68 | <.001 |
| | | Instructional limitation * Self-confidence | -0.54 | 2.46 | .828 |
| | | Class size * Self-confidence | -1.47 | 3.11 | .637 |
| | | School resources * Self-confidence | 0.68 | 2.78 | .809 |
| | | Valuing math | 16.46 | 2.94 | <.001 |
| | | Homework time | 5.35 | 2.22 | .016 |
| | | Home resources | 6.75 | 1.93 | .001 |
| | Random | $\tau_{00}$ | 8505.23 | 92.22 | <.001 |
| | | Extra lessons | 55.68 | 7.46 | .015 |
| | | Self-confidence | 76.88 | 8.77 | .023 |
| | | $\sigma^2$ | 2297.42 | 47.93 | |

As shown in Table 94, in comparing the amount of variance accounted for in five recent models, it seems that Model 22 worked more efficient than Models 9, 19, and 20, but less efficient than Model 21. Specifically, whereas the amount of explained variance between schools in Model 22 increased by 11% compared to Model 19, 12% compared to Model 9, and 13% compared to Model 21, it reduced by 2% when compared to Model 21. Thus, Model 21 serves as the best school background model to predict math achievement in South Africa.

Table 95.
*Comparison of* $R^2$ *between Model 22 and Previously Constructed Models 9 and 19-21 for South Africa*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 22 vs. 9 | 0.12 | 0.00 |
| 22 vs. 19 | 0.11 | 0.00 |
| 22 vs. 20 | 0.13 | 0.00 |
| 22 vs. 21 | -0.02 | 0.00 |

*Final Model*

With an intention to identify the most efficient and parsimonious model to predict eighth-grade math achievement in South Africa, Model 23 was built by including all the statistically significant level-2 predictors in earlier combined models to Model 9 and then compared with the three combined models, Models 14, 18, and 22. It is worth noting that in Model 23, the slope variance of self-confidence in learning math was set to 0 because it was not statistically significant in earlier combined models (i.e., Models 14 and 18).

Similar to Model 21, Model 23 with school resources as the only level-2 predictor in the model did not produce statistically significant cross-level interaction effects. All of the six remaining fixed effects, however, were statistically significant: (1) school resources ($\gamma$ = 58.23, *SE* = 26.31, *p* = .031), (2) extra math lessons ($\gamma$ = -8.52, *SE* = 2.40, *p* = .001), (3) student self-confidence in learning math ($\gamma$ = 23.65, *SE* = 3.18, *p* <.001), (4) student valuing of math ($\gamma$ = 16.32, *SE* = 2.81, *p* <.001), (5) time students spent on homework ($\gamma$ = 5.13, *SE* = 2.24, *p* = .022), and (6) home resources ($\gamma$ = 6.64, *SE* = 1.86, *p* <.001). Interestingly, in this model, the only slope that was allowed to vary (i.e., extra math lessons) was not statistically significant ($\gamma$ = 53.12, *SE* = 7.29, *p* = .07). These results suggest that, except for extra math lesson that consistently identified as having inverse relationship with math achievement, all other predictors in the model tended to have significantly positive relationship with math achievement in South Africa. School resources appeared to have the strongest positive relationship with math achievement. That is, a unit increased in school resources tended to be associated with an increased of 58.23 points in math achievement; whereas an unit increased in time students spent on homework was associated with only 5.13 points in math achievement.

Table 96.
*Parameter Estimates for Full Model for South Africa*

| Model | Type | Parameters | Estimates | *SE* | *p* |
|---|---|---|---|---|---|
| 23 | Fixed | INT | 162.13 | 17.95 | <.001 |
| | | School resources | 58.23 | 26.31 | .031 |
| | | Extra lessons | -8.52 | 2.40 | .001 |
| | | School resources * Extra lessons | -4.95 | 2.53 | .055 |
| | | Self-confidence | 23.65 | 3.18 | <.001 |
| | | Valuing math | 16.32 | 2.81 | <.001 |
| | | Homework time | 5.13 | 2.24 | .022 |
| | | Home resources | 6.64 | 1.86 | .001 |
| | Random | $\tau_{00}$ | 7598.68 | 87.17 | <.001 |
| | | Extra lessons | 53.12 | 7.29 | .066 |
| | | $\sigma^2$ | 2329.56 | 48.27 | |

Finally, as evident in Table 96, Model 23 appears to be the most efficient model for South Africa because this model accounted for the largest amount of variance between schools. Specifically, Model 23 accounted for 7% more of the between school variance when compared to Model 14, 11% when compared to Model 22 and 18% when compared to Model 28. Therefore, for South Africa, Model 23 serves as the best model for predicting math achievement.

Table 97.
*Comparison of* $R^2$ *between Model 22 and Previously Constructed Models 9 and 19-21 for South Africa*

| Compared Model | $\tau_{00}$ | $\sigma^2$ |
|---|---|---|
| 14 | 0.07 | -0.01 |
| 18 | 0.18 | -0.02 |
| 22 | 0.11 | -0.01 |

## Summary of Results

### *Missing Data*

In all four countries included in this study (i.e., Canada, USA, Egypt, and South Africa), missing data were present at both level-1 (i.e., student level) and level-2 (i.e., teacher and school level). With more missing data found at level-2 than at level-1, an evaluation of the extent to which these missing data at level-2 were completely at random was conducted for each country. The results showed that the missing data mechanism in all countries was not completely at random. In addition, the amount of missing data tended to be larger in developing countries (i.e., Egypt and South Africa) than in developed countries (i.e., Canada and USA). Using listwise deletion as a method of missing data treatment, the percentage of complete cases in Canada, USA, Egypt, and South Africa was 73.74%, 55.12%, 26.44%, and 17.56%, respectively, and the number of schools or level-2 units for these countries was 271, 153, 69, and 52, respectively.

### *Univariate Analysis*

Descriptively, considerable differences were observed in the weighted average math achievement across four countries, from 272.66 for South Africa to 529.30 for Canada. Similarly, it was observed that in Canada and USA, students had most basic resources at home for learning (i.e., calculator, desk, and computer) and rarely took extra math lessons but in Egypt and South Africa, students reported having fewer resources at home for learning and took extra math lessons more frequently, at least once or twice a week. However, there were also some commonalities across these countries. For example, students tended to spend a modest amount of time on math homework and had above medium level of self-confidence in learning math and valuing of math. In terms of opportunity to learn, number topic ranked first with 75.30% of students in South Africa, 95.85% of students in Canada, 99.13% of students in Egypt, and 99.65% of students in the U.S taught the subject before the TIMSS assessment. The areas that students had the least opportunity to learn appeared to be data (39.32% for South Africa and 61.02 for Egypt), algebra (59.11% for Canada) and geometry (70.10% for the U.S.).

At teacher and school level, it was observed that schools in Canada and U.S tended to have smaller class size than schools in Egypt and South Africa. Approximately 87% of the schools in Canada and 90% of the schools in the U.S. had less than 33 students in a class, whereas in Egypt and South Africa, the majority of schools (i.e., more than 70% in Egypt and 87% in South Africa) had between 33 and 41+ students in a class. Another interesting difference across the four countries was found in the average math instructional hours per year. Whereas schools in Canada and South Africa varied greatly in the average math instructional hours per year (i.e., 30 to 388 hours and 93 to 360 hours, respectively), there was a narrower range of average yearly math hours instructed in the U.S. and Egypt (i.e., 24.27 to 180 hours and 22.5 to 174.6, respectively). It is also worth noting that whereas 100% of Egyptian teachers reported being prepared to teach math content, less than half of teachers in Canada felt the same way. In the U.S. and South Africa, approximately two third of the teachers indicated they were prepared to teach math content. Similarities that were observed across these countries included teachers' participation in various math-related professional developments, their indication of readiness to teach math topics, the frequency of activities in math lessons, and the moderate amount of homework assigned to students.

### *Bivariate Analysis*

An examination of weighted bivariate correlations among student-level predictors (i.e., gender, student self-confidence in learning math, valuing of math, time students spent on homework, extra math lessons, and home resources) suggested that these predictors were uncorrelated with each other. From Egypt, South Africa to Canada and the U.S., weak correlation coefficients among these predictors were observed, less than +/-.40.

Similarly, the results of unweighted bivariate correlations among level-2 predictors suggested that these variables were weakly related to each other (*r* less than +/-.40), except for those that measure the same construct such as ready to teach math content variables, opportunity to learn variables, and activities in math lessons variables. In these instances, the correlation coefficients among the variables tended to be moderate in strength (*r* less than .70) and positive in direction.

### *Evaluation of HLM Assumptions*

For each of the four countries included in this study, visual analysis of both level-1 and level-2 random effects of the most efficient and parsimonious model was conducted. Results of these analyses suggested that in all four countries, the assumptions of normality and homogeneity of level-1 and level-2 random effects were satisfied.

### *HLM Analysis*

#### *Unconditional model*

The HLM analysis started with the unconditional model where none of the level-1 or level-2 predictor was included in the model. Across countries, the fixed effect for the intercept ranged from 267.57 (*SE* = 17.18) for South Africa to 527.33 (*SE* = 2.55) for

Canada. This suggests considerable differences in overall average school math scores across countries. Likewise, the intra-class correlation (ICC) was found dissimilar from one country to another, with Canada and Egypt having the lowest ICC of .28 and South Africa having the highest ICC of .76. These data suggest a modest to strong level of clustering of students occurred between schools across countries. In other words, approximately 28% to 76% of the total variance in math scores occurred between schools.

*Research Question 1*

To what extent are student background variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, and tutoring in math) associated with TIMSS 2003 eighth-grade math scores in each country?

In order to answer this research question, first, each of the student background variables was entered separately into the unconditional model to predict math achievement. Then, as a group, all of these variables were included in the combined student background model to predict math achievement. The results obtained from the combined student background models for Canada, USA, Egypt and South Africa suggest that this model worked differently across countries. Notably, in Canada, Egypt, and South Africa, all of the student background variables, except for gender, were statistically significant predictors of math achievement, whereas, in the U.S, only extra math lessons and student self-confidence in learning math showed statistically significant relationships with math achievement. Similarly, whereas the relationships between math achievement and valuing of math, and time students spent on homework appeared to differ significantly across schools in the U.S., these relationships did not seem to vary statistically across schools in Canada. Noteworthy was in Egypt where none of the slope variances was statistically significant, meaning that the relationships between math achievement and student background predictors did not differ significantly across schools.

However, there were some commonalities across these countries. For example, gender as a main effect did not seem to contribute significantly in the prediction of math in all of the four countries. Similarly, whereas student self-confidence in learning math was found to have the strongest and positive relationship with math achievement, extra math lessons tended to show an inverse relationship with math scores across countries. Put differently, across countries, the higher level of self-confidence students expressed in learning math, the higher math scores they tended to achieve. In contrast, the more frequently students took extra math lessons, the poorer in math they seemed to perform. Finally, the inclusion of student background variables in the unconditional model resulted in increased explained variance within schools in all of the countries although the amount of increase differed across countries.

*Research Question 2*

To what extent are home resources variables (i.e., availability of calculator, computer, and desk for student use) associated with TIMSS 2003 eighth-grade math scores in each country?

The results of the home resources model suggested that as a main effect, home resources was found to be a statistically significant predictor of math achievement in all countries. With the estimates for the fixed effect of home resources of 31.92 (*SE* = 3.65) for Egypt, 9.74 (*SE* = 1.86) for USA, 7.55 (*SE* = 2.30) for South Africa, and 7.48 (*SE* =

1.76) for Canada, it could be inferred that the more home resources students had for learning math, the higher math scores they tended to achieve. Interestingly, it was found that only in the U.S., the relationship between home resources and math achievement varied significantly across schools. Examining the model from the aspect of the variance accounted for it appeared that the use of this model instead of the unconditional model only yielded a marginal increase of the within school variance (from .4% in Canada to 5% in Egypt) but a considerable reduction of the between school variance in Canada (up to 56%).

As a next step of model building, all of the student variables (i.e., gender, self-confidence in learning math, valuing of math, time on math homework, tutoring in math and home resources) that were statistically significant in earlier models were included in the overall combined model or the foundational level-1 model to predict math achievement. Thus, as a result of this model building strategy, the foundational level-1 model was different across countries. For example, for Egypt, the foundational level-1 model was a random intercept only model, but for remaining countries, this model was a random coefficient model.

As different as countries could be, the results of the foundational level-1 model suggested that all of the level-1 predictors (i.e., self-confidence in learning math, valuing of math, time on math homework, tutoring in math and home resources) were statistically significant predictors of math achievement in Egypt and South Africa. For Canada, however, in the presence of other variables in the model, only extra math lessons, student self-confidence in learning math, and valuing of math showed statistically significant relationships with math achievement. In the U.S., in addition to extra math lessons and self-confidence in learning math, home resources also appeared statistically significant in the prediction of math achievement.

One thing that was similar across countries was that by using this model instead of the previous models, some considerable increases either in the within school variance or in the between school variance was observed. Specifically, the foundational level-1 model accounted for 37% more of the within school variance than the home resources model in Canada, 26% more of the within school variance than the home resources model in the U.S., 52% more of the between school variance than the student background model in Egypt, and 20% more of the within school variance than the home resources model in South Africa.

*Research Question 3*

To what extent are instructional variables (i.e., opportunity to learn, activities in math lessons, amount of homework assignment, and instructional time) associated with TIMSS 2003 eighth-grade math scores in each country?

When adding opportunity to learn math topics, activities in math lessons, amount of homework assignment, and average math instructional hours per year as level-2 predictors to the foundational level-1 model to predict math achievement, the combined instructional practice model produced interesting results across countries. In Canada, this model yielded five statistically significant cross-level interactions effects: (1) average math instructional hours per year by gender, (2) opportunity to learn algebra by extra math lessons, (3) opportunity to learn geometry by extra math lessons, (4) opportunity to learn data by self-confidence in learning math, and finally, (5) opportunity to learn measurement by self-confidence in learning math. The nature of these interactions was detailed in the HLM analysis for Canada.

In the U.S., the combined instructional practice model produced three statistically significant cross-level interaction effects: (1) opportunity to learn data by self-confidence in learning math, (2) opportunity to learn measurement by student valuing of math, and (3) opportunity to learn geometry by time student spent on homework. Again, the nature of these interactions can be found in the HLM analysis for USA. For Egypt and South Africa, however, no statistically significant cross-level interaction effect was detected. In fact, in both countries, none of the level-2 main effects was statistically significant, either. In sum, the use of this model instead of the foundational level-1 model resulted in an increase in the explained variance between schools by 25% for the U.S, 30% for Canada, 6% for Egypt, and 16% for South Africa.

*Research Question 4*

To what extent are teacher-related variables (i.e., preparation to teach, ready to teach, and professional development) associated with TIMSS 2003 eighth-grade math scores in each country?

The combined teacher-background model with preparation to teach math content, ready to teach math topics and math-related professional development as level-2 predictors also yielded interesting results across countries. Specifically, in Canada, two statistically significant cross-level interaction effects were detected: student self-confidence in learning math by preparation to teach math content and student self-confidence in learning math by math-related professional development. The nature of these interactions can be found in the HLM analysis for Canada. As for the U.S., one statistically significant cross-level interaction effect was observed between ready to teach number and time students spent on homework. Details of this interaction were illustrated in the HLM analysis for USA. In Egypt, math-related professional development was found to be the only level-2 predictor that statistically significantly related to math achievement. Finally, for South Africa, neither cross-level interaction effects nor level-2 main effects for teacher-background variables was found statistically significant.

In terms of model fit, the use of this model instead of the foundational level-1 model resulted in an increase in the explained variance between schools by 2% for the U.S, 21% for Canada, 11% for Egypt, and 4% for South Africa.

*Research Question 5*

To what extent are school-related variables (i.e., class size, school resources for math instruction, and teacher perception of math instructional limitations due to student factors) associated with TIMSS 2003 eighth-grade math scores in each country?

The combined school-background model with class size, school resources for math instruction, and teacher perception of math instructional limitations due to student factors as level-2 predictors produced different results across countries. Specifically, in Canada, no statistically significant cross-level interaction effect was detected but there were two statistically significant level-2 main effects: class size for math instruction and teacher perception of math instructional limitations due to student factors. In the U.S., however, two statistically significant cross-level interaction effects were observed. One was between class size for math instruction and student self-confidence in learning math and the other interaction was between class size for math instruction and student valuing of math. Interestingly, for Egypt, no statistically significant level-2 main effect was found. As for South Africa, school resources appeared to contribute statistically significantly to the prediction of math achievement.

An evaluation of the proportion of variance accounted for between schools suggested that, compared to the foundational level-1 model, the combined school-background model was less efficient for USA, equally efficient for Egypt, and more efficient for Canada and South Africa. Specifically, a reduction of 7% of the explained between school variance was noted for USA, whereas an increase of 16% was noted for Canada and 12% for South Africa.

*Final Model*

With an intention to identify the most efficient and parsimonious model to predict eighth-grade math achievement in each of the country, the final model was built and compared with the three combined models (i.e., instructional practice model, teacher-background model, and school-background model). It is also worth noting that this model included only fixed and random effects that were statistically significant in earlier combined models.

In Canada, this final model produced six statistically significant cross-level interaction effects: (1) average math instructional hours per year by gender, (2) opportunity to learn data by gender, (3) opportunity to learn algebra by extra math lessons, (4) opportunity to learn geometry by extra math lessons, (5) opportunity to learn data by self-confidence in learning math, and (6) preparation to teach math content by self-confidence in learning math. Details of these interactions can be found in the HLM analysis for Canada. In the U.S., two statistically significant cross-level interaction effects were observed: opportunity to learn geometry by time student spent on homework and ready to teach number by time student spent on homework. Again, the nature of these interactions was discussed in detail in the HLM analysis for USA. In Egypt, the only level-2 predictor, math-related professional development that was included in the final model showed statistically significant relationship with math achievement. Similarly, for South Africa, the only level-2 predictor, school resources that was included in the final model showed statistically significant contribution to the prediction of student math scores.

In comparison with previously constructed combined models (i.e., instructional practice model, teacher-background model, and school-background model), it appeared that the final model served as the best model for predicting math achievement in Canada and South Africa; whereas for USA, the combined instructional practice model worked the best, and for Egypt, the combined teacher-background model served as the most efficient and parsimonious model for predicting math achievement.

# CHAPTER FIVE

# DISCUSSION

## Purpose

The purpose of this study was to investigate correlates of math achievement in both developed and developing countries. Specifically, two developed countries, the United States and Canada and two developing countries, Egypt and South Africa that participated in the TIMSS 2003 eighth-grade math assessment were selected for this study. For each country, a series of two-level models were constructed to examine the extent to which student background, home resources, instructional practices, teacher background, and school background-related variables were associated with TIMSS 2003 eighth-grade math scores in the respective country. Ultimately, the overall goal for this study was to provide empirical evidence that supports the rationale that developing countries should not implement educational models that appear to work in developed countries; rather, they should develop and implement their own educational models based upon their countries' specific research findings. This is because countries differ in characteristics and a model that works in a developed country might not work in a developing country (Bryan et al., 2007; Delaney, 2000; Watkins & Biggs, 2001).

## Review of Method

This study used secondary data from the TIMSS 2003 to investigate the relationships between eighth-grade student math achievement and contextual and background factors in four countries, the United States, Canada, Egypt, and South Africa.

The dependent variable of the study was overall math score, an IRT-based score, which was calculated by averaging five plausible sub-topic scores: algebra, number, geometry, measurement, and data. The independent variables were five groups of factors: (1) student background, (2) home resources, (3) instructional practices, (4) teacher background, and (5) school background. Each of these groups of factors was precisely defined by using existing variables in the TIMSS 2003 database. The study's theoretical framework, Carroll's (1963) model of school learning, as well as a review of related literature guided the selection of these variables. In addition, subject matter experts were consulted and factor analysis was conducted to provide reliability evidence for the variables included in this study.

Because of the naturally occurring clusters in the TIMSS 2003 data, multilevel models were used to capture the relationships among the student level and the teacher/school level variables and eighth-grade student math achievement. Specifically, for each country, 23 models were constructed to represent the student level (level-1) and the teacher/school level (level-2) in the TIMSS 2003 data. The HLM analysis started with an unconditional or baseline model where none of the level-1 or level-2 variables was included. Next, each student background variable was entered separately in the baseline model to make Models 2-6. Model 7 was constructed with all the student background variables included in the baseline model. Next, home resources was added to the baseline model to make Model 8. Then, all of the variables that were statistically significant in Models 7 and 8 were included to the baseline model to make the foundational student model or Model 9.

At level-2, Models 10-13 were built by first entering individual instructional practice variable to the foundational level-1 model. Then, as a group, all instructional practices variables were added to the foundational level-1 model to make the combined instructional practice model (Model 14). Likewise, teacher background models (Models 15-18) and school-background models (Models 19-21) were constructed. Finally, Model 23 was built by adding all level-2 variables that were statistically significant in the earlier combined models (Models 14, 18, and 22) to the foundational level-1 model.

Prior to analysis of the data, listwise deletion as a missing data treatment method was employed to eliminate all the missing data at both the student level and the classroom/school level. This step was conducted because in two-level HLM analyses, parameter estimates are computed based on complete cases. In addition, because TIMSS 2003 utilized a complex sampling design, a normalized student sampling weight was used in all analyses of the data in order to obtain more accurate population estimates.

## Results

### *Unconditional model*

The HLM analysis started with the unconditional model where none of the level-1 or level-2 predictor was included. Results from this model suggest that considerable differences in overall average school math scores existed across countries, from 267.57 (*SE* = 17.18) for South Africa to 527.33 (*SE* = 2.55) for Canada. Likewise, the intra-class correlation (ICC) was found dissimilar from one country to another. Whereas the ICCs were relatively small (.28) for Canada and Egypt, they were relatively large for the United States (.51) and South Africa (.76). In other words, approximately 28%-76% of the variance in math achievement occurred between schools in these countries.

One possible explanation for a high level of ICC as observed in the United States and South Africa could be the sampling procedures implemented by the TIMSS 2003 (i.e., in each school, only one math class was sampled). Because students from the same school had similar opportunities to learn math and were taught by the same math teacher, their math scores were more homogeneous when compared within schools than between schools. As a result, it was more likely to find statistically significant effects for level-2 contextual and background factors in subsequent models (i.e., instructional practices, teacher background, and school background models).

*Student Background Model*

The student background model was developed to address the first research question regarding the extent to which eighth-grade math achievement was associated with gender, self-confidence in learning math, valuing of math, time students spent on homework, and tutoring in math in the countries of Canada, USA, Egypt and South Africa. The results suggested that this model worked differently across countries. Specifically, in Canada, Egypt, and South Africa, all of the student background variables, except for gender, were statistically significant predictors of math achievement; whereas, in the U.S, only extra math lessons and student self-confidence in learning math showed statistically significant relationships with math achievement. Similarly, whereas the relationships between math achievement and valuing of math, and time students spent on homework appeared to differ significantly across schools in the U.S., these relationships did not seem to vary statistically across schools in Canada. Interestingly, in Egypt, none of the relationships between math achievement and student background predictors appeared to differ significantly across schools.

In connecting to existing literature (Beaton et al., 1996; Mullis et al., 2000; Peterson & Fennema,1985; Rodriguez, 2004), these results do not support the view that a gender gap exists in math achievement. However, this study concurs with prior research findings (House, 2006; Koller, Baumert & Schnabel, 2001; Pajeres & Graham, 1999) in that student self-confidence in learning math contributes positively and significantly in the prediction of student math achievement. In fact, in this study, evidence was found in all four countries that, of all student background variables, student self-confidence in learning math showed the strongest and positive relationship with math achievement. As for tutoring/extra math lessons, consistent results were found between this study and those of Papanastasiou (2002) that the more frequently students took extra math lessons, the poorer in math they seemed to perform. It is worth noting that Papanastasiou (2002) looked at fourth-grade student in Cyprus, Hong Kong and the United States; whereas this study focused on eighth-grade students in Canada, the United States, Egypt and South Africa.

*Home Resources Model*

The home resources model aimed at answering the second research question regarding the association between eighth-grade math achievement and the availability of a calculator, desk, and computer for student use at home. Without controlling for other variables, home resources as a sum composite variable of calculator, desk and computer was found to be a statistically significant predictor of math achievement in all four countries of the study Canada, USA, Egypt and South Africa. Specifically, in these countries, the more home resources students had for learning math, the higher math scores students tended to achieve.

Interestingly, however, after controlling for other student background-related variables in the foundational level-1 model, home resources no longer was a statistically significant predictor of math achievement in Canada. In other words, in the presence of other student background variables, home resources remained as a potential predictor of eighth-grade math achievement in three of the four countries, the United States, Egypt, and South Africa. These results were congruent with the findings from the recent study of Mullis, Martin, Gonzalez and Chrostowski (2004). According to these researchers (Mullis et al., 2004), students from homes with a range of study aids such as computer, calculator, desk, and dictionary tended to have higher math scores than their peers who did not have access to such resources at home.

*Instructional Practices Model*

The instructional practices model focused on the third research question concerning the relationships between eighth-grade math achievement and level-2 instructional practices-related predictors such as opportunity to learn math topics, activities in math lessons, amount of homework assignment, and average math instructional hours per year. This model produced interesting results across countries. Specifically, whereas there were five statistically significant cross-level interactions effects in Canada and three in the United States, none was detected in Egypt and South Africa. In fact, in these two developing countries, none of the level-2 main effects was statistically significant, either.

Unexpectedly, in Canada, it was found that the more math instructional hours the students had, the poorer math scores they tended to achieve. Although this pattern was observed for both gender groups, the negative effect seemed stronger for male than female students. Similarly, increases in opportunity to learn geometry were associated with lower math scores for students who reported taking extra math lessons sometimes to almost everyday. Likewise, increases in opportunity to learn data were related to poorer math achievement for all students, especially those with low self-confidence in learning math. However, the results also showed that with more opportunity to learn measurement, students with a higher level of self-confidence in learning math tended to perform better in math than their peers who had a lower level of self-confidence in learning math. Similarly, increases in opportunity to learn algebra were associated with better math performance for those students who tended to take extra math lessons sometimes to almost everyday.

Interestingly, in the United States, whereas opportunity to learn measurement was found to have positive relationship with math achievement and this relationship was observed for all students, regardless of their levels of valuing math, opportunity to learn data was found to be inversely associated with math achievement and this relationship was observed for all students, regardless of their levels of self-confidence in learning math. Also, it was surprising to find that, for students who reported spending little time on homework, increases in opportunity to learn geometry was associated with higher math scores. However, for students who reported spending a medium to high amount of time on homework, the higher opportunity to learn geometry was related to poorer math scores.

Clearly, these results showed consensus with existing literature (Baker, 1993; Westbury, 1992; Wiley &Yoon, 1995) that opportunity to learn math was associated with math achievement. Although it appears that different math topics (i.e., number, algebra, measurement, geometry, and data) were related to math achievement in different ways, the finding of this study were supported by Wiley and Yoon (1995). Specifically, Wiley and Yoon (1995) concluded that students' exposure to different math topics, and the way in which these topics were covered affected students' performance on tests.

*Teacher Background Model*

The fourth research question was addressed by the teacher background model which centered on the relationship between math achievement and three level-2 teacher background-related variables: preparation to teach math content, ready to teach math topics, and math-related professional development. This model yielded interesting findings across countries. Whereas two statistically significant cross-level interactions were detected for Canada, only one was observed for the United States. Similarly, whereas math-related professional development was found significantly and positively related to math achievement in Egypt, neither cross-level interaction effects nor level-2 main effects for teacher-background variables were found statistically significant in South Africa.

As expected, in Canada, students whose teachers reported being prepared to teach math content consistently performed higher in math than their peers whose teachers reported being not prepared to teach math content. The achievement gap among these students, however, was large when the students expressed low self-confidence in learning math, and became narrower as their self-confidence in learning math increased. Interestingly, the data in Canada also showed that there was an inverse relationship between math achievement and math-related professional development. That is, teachers' participation in more math-related professional development programs did not result in higher math performance for their students, regardless of how self-confident they were in learning math However, among the students, those who had a higher level of self-confidence in learning math consistently outperformed their peers who had a lower level of self-confidence in learning math. The differences in their math achievement appeared to be largest when their teachers had five professional development programs and smallest when their teachers had none of these programs.

Unexpectedly, the data for the United States suggested that students whose teachers were very ready to teach number tended to achieve lower math scores than their peers whose teachers were ready to teach number. The achievement gap between these students, however, was small when their time on homework was low and became more substantial when the amount of time they spent on homework was high. These results should be interpreted with cautions because they were based on 149 teachers who reported very ready to teach and only 4 teachers who reported ready to teach.

In linking with existing literature, the observed positive relationship between math achievement and preparation to teach math content in Canada appeared to be congruent with evidence from recent research (Greenberg, Rhodes, Ye, & Stancavage, 2004; Grouws, Smith, & Sztajn, 2004). Using the data from NAEP 2000 for eighth grade math, Greenberg, Rhodes, Ye, and Stancavage (2004) showed that students across all math ability levels (i.e., low, medium, and high) who had teachers with a major in math scored higher than their peers whose teachers had a major outside of their field of teaching.

Likewise, the results observed for Canada and South Africa with regard to the relationship between math achievement and math-related professional development were found consistent with the findings from prior research. Specifically, Wiley and Yoon (1995) examined the data from the California Learning Assessment System (CLAS) 1993 and found that Grades 4 and 8 students whose teachers were familiar with mathematics instruction assessment guides and participated in mathematics curriculum activities tended to perform significantly better than their peers whose teachers were not involved in those activities. For Grade 10 math achievement, however, little impact was noted despite the highest level of teachers' familiarity with math goals and standards and frequent participation in various instructional activities.

*School Background Model*

The school background model focused on the last research question that examined the relationship between math achievement and level-2 school-related factors: class size, school resources for math instruction, and teacher perception of math instructional limitations due to student factors. The results suggested that whereas class size for math instruction and teacher perception of math instructional limitations due to student factors were statistically significant predictors of math achievement in Canada and school resources was a significant predictor of math achievement in South Africa, none of level-2 main effects was statistically significant for Egypt.

In the United States, however, two statistically significant cross-level interaction effects were observed. The nature of the interaction between class size for math instruction and student self-confidence in learning math indicated that for students who reported having high self-confidence in learning math, changes from small class size (i.e., 1-24 students) to large class size (i.e., 41+ students) tended to lower their math scores significantly. Conversely, for students who reported having low self-confidence in learning math, increases in class size appeared to improve their math scores. Thus, math achievement gap among eighth-grade students was most substantial when they learned math in small class size and became smaller when they learned math in large class size.

As for the interaction between class size for math instruction and student valuing of math, the data suggested that in schools with small class size (i.e., 1-24 students), students who reported having a low level of valuing math tended to achieve higher math scores than their peers who reported having medium or high levels of valuing math. This pattern of relationship, however, appears to reverse in schools with larger classes. That is, students with medium and high levels of valuing math tended to perform better than their peers who reported having a low level of valuing math. Nevertheless, a similar trend was noted for all of the students, regardless of their levels of valuing math. That is, changes from smaller classes to big classes were associated with increased math scores.

There appear some agreements between these results and those of prior research. For example, Baker et al. (2002) concluded that the effect of school resources on achievement within nations was not associated with national income levels. In this study, although school resources was positively related to math achievement in South Africa, it was not found statistically significant in Egypt. Similarly, the positive association found between class size and math achievement in Canada and the interesting interaction patterns between class size and math achievement in relation to student self-confidence in learning math and student valuing of math in the United States support the mixed results suggested by existing literature about the relationship between class size and student learning (Cooper, 1989b; ERS, 1978; Nye, Hedges, & Konstantopoulos, 2001; Pong & Pallas, 2001; Woobmann & West, 2006). Specifically, Pong and Pallas (2001) and

Woobmann and West (2006) found that in Canada and some other European countries such as Greece and Iceland, schools with larger classes tended to be associated with better performance in math than did schools with smaller classes. In contrast, the results from a longitudinal study conducted by Nye, Hedges, and Konstantopoulos (2001) in the United States suggested that the benefits of small class size in terms of student achievement persisted significantly throughout the six years of the study.

### *Final Model*

With an intention to identify the most efficient and parsimonious model to predict eighth-grade math achievement in four countries examined in this study, the final model was built and compared with the three models: instructional practices model, teacher-background model, and school-background model. It is worth noting that the final model included only fixed and random effects that were statistically significant in earlier models. The results of these comparisons suggested that the final model served as the best model for predicting math achievement in Canada and South Africa; whereas for the USA, the instructional practices model worked the best, and for Egypt, the combined teacher-background model served as the most efficient and parsimonious model for predicting math achievement.

## Limitations

Findings of this study must be interpreted in light of its limitations. First, the massive amount of missing data (i.e., from 26.26% for Canada to 82.44% in South Africa) due to sampling procedures (i.e., multistage, stratified, and unequal probability), assessment design (each student took only one test booklet or a subset of the entire test items), and non-responses from participants could negatively affect accuracy of this study results, especially when the missing data mechanism in each country was found not completely at random. Across the four countries, it was interesting to observe that the variable average math instructional hours per year presented the majority of missing data. The amount of missingness for this variable alone was 19% for Canada, 31% for the United States, 61% for Egypt, and 70% for South Africa. It is important to note that this variable was created by the TIMSS 2003 using items from both the teacher survey and the school survey (see Table 6 for further details). It could be possible that the way these items were designed and administered was associated with the amount of their missingness. Thus, it is worthwhile for future TIMSS studies to revisit the design of these items as well as the administration of these surveys in order to maximize participants' responses.

Second, because this is an analysis of secondary data, the study was limited to the data collected in the TIMSS 2003. As an example, the construct of home resources in this study was originally conceptualized to include four indicators: the availability of a calculator, dictionary, desk and computer for student use at home. Interestingly, it was found that whereas the data for the variable dictionary were available for Canada, Egypt, and South Africa, they were not available for USA. Specifically, all of the students in USA had missing data on this variable. In an attempt to understand why such data were not available for USA, the researcher contacted the TIMSS 2003 Office. Because further investigation of the problem was needed at the TIMSS 2003 Office, a decision was made to recalculate the composite variable home resources to include only three indicators: calculator, desk, and computer.

Also, although the TIMSS 2003 provides rich background and contextual information, there were variables that the researcher wished to have but the TIMSS 2003 database did not have. For example, there was no measure of student self-confidence in learning individual math topics (i.e., number, algebra, measurement, geometry and data) or students' past achievement or aptitude scores. For this reason, it was not possible to establish a direct connection between the variables proposed by Caroll's (1963) model of school learning and the variables selected for this study.

Third, the results of this study were based on the relationships between student math achievement and contextual and background factors which were self-reported by students, teachers, and school principals. Self-reported data, according to Rosenberg, Greenfield, and Dimick (2006), have several potential sources of bias such as selective memory (remembering or not remembering experiences or events that occurred sometime in the past), telescoping (recalling events that occurred at one time as if they had occurred at another time), and social desirability (reporting behaviors that tend to be widely accepted by certain social groups rather than the behaviors actually exhibited by the respondents). Thus, it is important to interpret findings of this study with this limitation in mind.

Last but not least, analysis of secondary data with a complex survey design often requires the use of sampling weights. However, at the time this dissertation was conducted, common statistical software such as SAS and SPSS did not have the ability to incorporate sampling weights into multilevel analysis. Although the latest HLM version 6.06 was able to apply sampling weights, some parts of the HLM analysis output (i.e., table of fixed effects with regular standard errors) were not produced. Thus, in this study, tables of fixed effects with robust standard errors were reported. Additionally, as part of the HLM analysis, reliability estimates of the level-1 random coefficients were computed. However, reliability estimates of the random slopes tended to be lower than those of the intercepts. According to Raudenbush, Bryk, Cheong, and Congdon (2004), “Low reliabilities do not invalidate the HLM analysis. Very low reliabilities (e.g., <.10), often indicate that a random coefficient might be considered fixed in subsequent analyses.” (p.82).

## Implications

Despite its limitation, this study contributes significantly to the field of educational research. First, this study made an attempt to reduce bias in international educational research by examining correlates of math achievement in both developed and developing countries. Second, by investigating correlates of math achievement within each country rather than between countries, this study produced country-specific research findings related to eighth-grade students’ math achievement. For the national leaders, policy makers, and educators from these countries, especially developing countries, the results of this study could be insightful because they were carefully examined while controlling for the uniqueness of each country (i.e., student background, home resources, instructional practices, teacher background, and school resources). Thus, in each of the four countries, these results could be used directly to help evaluate and improve the effectiveness of their current mathematics educational systems.

Third, findings of this study provide strong evidence to support the view that countries differ and an educational model that worked for a developed country might not work for a developing country. Fourth, with detailed descriptions of research design, method and data analysis steps, this study serves as an example for other researchers to replicate this study either with other countries that participated in the TIMSS 2003 assessment or with other international achievement databases. Finally, by outlining limitations with the TIMSS 2003 data, this study aimed to provide TIMSS researchers with suggestions for refinement and improvement of future TIMSS studies.

## Future Research

As a result of this work, a number of future research studies can be conducted. First of all, in an effort to reduce bias in international achievement research, this study can be replicated with other developed and developing countries that participated in the TIMSS 2003 study. Ideally, the new studies should include countries from different continents in order to maximize variances across countries. Second, future work can be conducted using different existing large-scale international achievement data such as PIRLS and PISA. Because different databases tended to provide different contextual and background variables, it will be interesting to find out whether similar country-specific findings will result from the use of similar models with similar composite variables but different indicators. Third, it may also be worthwhile to consider building country-specific achievement models for different subjects such as Science and Reading, and for different grades such as Fourth, Eighth, and Twelfth in future research. Finally, because the current study did not explain why certain relationships between math achievement and contextual and background factors were present or absent, further studies can be conducted within each country to gain deeper understanding of the reasons underlying these relationships.

REFERENCES

Abu-Hilal, M. M. (2000). A structural model of attitudes toward school subjects, academic aspirations, and achievement. *Educational Psychology, 20,* 75–84.

Adler, C. (1996). *Report of the public committee on long school day*. Jerusalem: Center for the Study of Social Policy in Israel.

Akiba, M., LeTendre, G. K., & Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher, 36*(7), 369-387.

Allison, P.D. (2001) *Missing data*. Thousand Oaks,CA: Sage.

Allinder, R. M. (1994). The relationship between efficacy and the instructional practices of special education teachers and consultants. *Teacher Education and Special Education, 17*, 86-95.

Angrist, J. D., & Lavy, V. (2001). Does teacher training affect pupil learning? Evidence from matched comparisons in Jerusalem public schools. *Journal of Labor Economics, 19* (2), *343-369.*

Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement.* New York: Longman.

Baker, D. P. (1993). A rejoinder. *Educational Researcher, 22* (3), 25-26

Baker, D. P., Fabrega, R., Galindo, C., & Mishook, J. (2004). Instructional time and national achievement: Cross-national evidence. *Prospect, 34* (3), 311-334.

Baker, D. P., Goesling, B., & LeTendre, G. K. (2002). Socioeconomic status, school quality, and national economic development: A cross-national analysis of the "Heyneman-Loxley effect" on mathematics and science achievement. *Comparative Education Review, 46* (3), 291-312.

Baker, D. P., & LeTendre, G. K. (2005). *National differences, global similarities – world culture and the future of schooling*. Stanford, CA: Stanford University Press.

Bankov, K., Mikova, D., & Smith, T. M. (2006). School quality and equity in central and eastern Europe: Assessing between-school variation in educational resources and mathematics and science achievement in bulgaria. *Prospects, 36* (4), 448-473.

Beaton, A. E. (1998). Comparing cross-national student performance on TIMSS using different test items. *International Journal of Educational Research, 29,* 529-542.

Beaton, A. E., Mullis, I. V. S., Martin, M. O. Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA's third international mathematics and science study (TIMSS).* Chestnut Hill, MA: Boston College.

Beck, E. L. (1999). Prevention and intervention programming: Lessons from an after-school program. *Urban Review, 31* (1), 107-124.

Bielinski, J., & Davison, M. L. (2001). A sex difference by item difficulty interaction in multiple-choice mathematics items administered to national probability samples. *Journal of Educational Measurement*, *38*, 51-77.

Bolger, N., & Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement. *Journal of Educational Measurement, 27* (2), 165–174.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*, (8), 3-15.

Borko, H., & Putnam, R. (1996). *Learning to teach*. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (673-708). New York: Macmillan.

Bryan, C. A., Wang, T., Perry, B., Wong, N., & Cai, J. (2007). Comparison and contrast: Similarities and differences of teachers' views of effective mathematics teaching and learning from four regions. *ZDM Mathematics Education, 39*, 329-340.

Bush, G. W. (2001). President honors nation's leading math and science teachers. Retrieved on September 20, 2007 from http://www.whitehouse.gov/news/releases/2001/03/20010305-12.html.

Carpenter, T. P., & Fennema, E. (1992). Cognitively guided instruction: Building on the knowledge of students and teachers. In W. Secada (Ed.), *Researching educational reform: The case of school mathematics in the United States* (pp. 457-470). Special issue of International Journal of Educational Research.

Carpenter, T. P., Blanton, M. L., Cobb, P., Franke, M. L., Kaput, J., & McClain, K. (2004). *Scaling up innovative practices in mathematics and science.* Madison: University of Wisconsin-Madison, National Center for Improving Student Learning and Achievement in Mathematics and Science.

Carroll, J. B. 1963. A model of school learning. *Teachers College Record*, *64*, 722–733.

Carter, D. S. G., & O'Neill, M. H. (1995). *International perspectives on educational reform and policy implementation*. Washington, D.C.: Falmer Press.

Cogan S. L., & Schmidt, W. H. (1999). An examination of instructional practices in six countries. In G. Kaiser, E. Luna, & I. Huntley (Eds.), *International comparisons in mathematics education* (pp. 68-85). Philadelphia, PA: Falmer Press.

Cohen, D. K., & Hill, H. C. (1998). *Instructional policy and classroom performance: The mathematics reform in California (RR-39).* Philadelphia: Consortium for Policy Research in Education.

Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record, 102* (2), 294-343.

Coladarci, T. (1992). Teachers' sense of efficacy and commitment to teaching. *Journal of Experimental Education, 60*, 323-337.

Coleman, J. S. (1975). Methods and results in the IEA studies of effects of school on learning. *Review of Educational Research, 45*(3), 355-386.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M.,Weinfeld, F. D.,& York, R. L. (1966). *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office.

Comber, L. C., & Keeves, J. P. (1973). Science education in nineteen countries. *International studies in evaluation*, Vol. 1. Stockholm: Almqvist & Wiksell.

Cooper, H. M. (1989a). *Homework.* White Plains, NY: Longman

Cooper, H. M. (1989b). Does reducing student-to-instructor ratios affect achievement? *Educational Psychologist, 24* (1), 78-98.

Cooper, H., Lindsay, J. J., Nye, B., & Greathouse, S. (1998). Relationships among attitudes about homework, amount of homework assigned and completed, and student achievement. *Journal of Educational Psychology, 90* (1), 70–83.

Cooper, H., & Valentine, J. C. (2001). Using research to answer practical questions about homework. *Educational Psychologist, 36* (3), 143-153.

Corcoran, T. B. (1995). *Transforming professional development for teachers: A guide for state policymakers.* Washington, DC: National Governors' Association.

Cosden, M., Morrison, G., Albanese, A. L., & Macias, S. (2001). When homework is not home work: After-school programs for homework assistance. *Educational Psychologist, 36* (3), 211-221.

Darling-Hammond, L. (2000). How teacher education matters. *Journal of Teacher Education, 51* (3), 166-173.

Wiley, D. E. & Yoon, B. (1995). *Educational Evaluation and Policy Analysis*, 1*7* (3), 355-370.

Davis, H., & Carr, M. (2001). Gender differences in mathematics: Strategy, use, the influence of temperament. *Learning and Individual Differences, 13,* 83-95.

Delaney, P. (2000). Study finds Asian countries are best in math, science: Newest TIMSS data indicates little progress for US 8th graders. *The Boston College Chronicle, 9* (8). Retrieved June 15, 2007 from http://www.bc.edu/bc_org/rvp/pubaf/chronicle/v9/d14/timss.html.

De Jong, R., Westerhof, K. J., & Creemers, B. P. M. (2000). Homework and student math achievement in junior high schools. *Educational Research and Evaluation, 6*, 130-157.

Educational Research Service. (1978). *Class size: A summary of research.* Arlington, VA: Educational Research Service.

Educational Research Service. (1980). *Class size research: A critique of recent meta-analyses.* Arlington, VA: Educational Research Service.

Eklof, H. (2007). Test-taking motivation and mathematics performance in TIMSS 2003. *International Journal of Testing, 7* (3), 311-326.

Encyclopedia of Educational Research (1960). London: Collier-Macmillan Limited.

Ercikan, K., McCreith, T. & Lapointe, T. (2005). *How are non-school related factors associated with participation and achievement in science? An examination of gender differences in Canada, the USA and Norway*. In S. J. Howie and T. Plomp (Eds), Contexts of learning mathematics and science: Lessons learned from TIMSS, 211-225. Netherlands: Swets & Zeitlinger International Publishers.

Evertson, C. (1980). Differences in instruction activities in high and low achieving junior high classes. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, *28* (2), 465-498.

Fennema, E., Carpenter, T. P., Jacobs, V. R., Franke, M. L., & Levi, L. W. (1998). A longitudinal study of gender differences in young children's mathematical thinking. *Educational Researcher*, *27* (5), 6-11.

Forshay, A. W., Thorndike, R. L., Hoytat, F., Pidgeon, D. A., & Walker, D. A. (1962). Educational achievement of thirteen-year-olds. Hamburg: UNESCO Institute for Education.

Frederick, W. C., & Walberg, H. J. (1980). Learning as a function of time. *Journal of Educational Research, 73*, 183–194.

Fuller, B. (1987). What school factors raise achievement in the third world? *Review of Educational Research. 57* (3), 255-292.

Glazerman, S., Mayer, D., & Decker, P. (2006). Alternative routers to teaching: the impacts of teach for America on student achievement and other outcomes. *Journal of Policy Analysis and Management, 25* (1), 75-96.

Glass, G. V., & Smith, M. L. (1978). *Meta-analysis of research on the relationship of class-size and achievement*. San Francisco: Far West Laboratory for Educational Research and Development.

Gonzales, P. (2001). *Using TIMSS to analyze correlates of performance variation in mathematics.* Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.

Goals 2000. (1994). Goals 2000: Educate America Act of 1994. Pub. L. No. 103-227, Stat. 125, 108.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76* (4), 569-582.

Greenberg, E., Rhodes, D., Ye, X., & Stancavage, F. (2004). *Prepared to teach: Teacher preparation and student achievement in eighth-grade mathematics*: American Institutes for Research. Paper presented at AERA 2004, San Diego, CA. Downloaded October 10, 2007 from http://www.air.org/news_events/documents/AERA2004PreparedtoTeach.pdf

Guskey, T. R. (1987). Context variables that affect measures of teacher efficacy. *Journal of Educational Research, 81* (1), 41-47.

Grouws, D. Smith, M., & Sztajn, P. (2004). The preparation and teaching practices of United States mathematics teachers: Grades 4 and 8. In P. Kloosterman & F. Lester, Jr. (Eds.), *Results and interpretations of the 1990 through 2000 mathematics assessments of the National Assessment of Educational Progress* (pp. 221-267). Reston, VA: National Council of Teachers of Mathematics.

Hahs-Vaughn, D. L. (2005). A primer for using and understanding weights with national datasets. *Journal of Experimental Education, 73,* 221-228.

Hambleton, R.K. Merenda, P.F. & Spielberger, C.D. (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: L. Erlbaum Associates.

Hamilton, L. S., McCaffrey, D. F., Stecher, B. M., Klein, S. P., Robyn, A., & Bugliari, D. (2003). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Educational Evaluation and Policy Analysis, 25* (1), 1-29.

Harris, A. M., & Carlton, S. T. (1993). Patterns of gender differences on mathematics items on the Scholastic Aptitude Test. *Applied Measurement in Education, 6,* 137-151.

Heyneman, S. P., & Loxley, W. A. (1982). Influences on academic achievement across high and low income countries: A re-analysis of IEA data. *Sociology of Education, 55* (1), 13-21.

Heyneman, S. P. & Loxley, W. A. (1983). The effect of primary-school quality on academic achievement across twenty-nine high- and low-income countries. *American Journal of Sociology, 88* (6), 1162-1194.

Hiebert, J., & Stigler, J. W. (2000). A proposal for improving classroom teaching: Lessons from the TIMSS video study. *Elementary School Journal, 1001,* 3-20.

Hilton, T. L. (1992). *Using national data bases in educational research*. Hillsdale, NJ: Lawrence Earlbaum.

Hofferth, S.L. (2005). Secondary data analysis in family research. *Journal of Marriage and Family, 67*, 891-907.

Hollins, E. R. (1995). Revealing the deep meaning of culture in school learning: Framing a new paradigm for teacher preparation. *Action in Teacher Education, 17* (1), 70-79.

House, J. D. (2006). Mathematics beliefs and achievement of elementary school students in Japan and the United States: Results from the third international mathematics and science study, *The Journal of Genetic Psychology, 2006, 167* (1), 31–45.

Hox, J. J. (1995). *Applied multi-level analysis, 2nd Ed.* Amsterdam: TT-Publikaties.

Huffman, D., Thomas, K. & Lawrenz, F. (2003). Relationship between professional development, teachers' instructional practices, and the achievement of students in science and mathematics. *School Science and Mathematics*, *103* (8), 378-387

International Association for the Evaluation of Educational Achievement. (2007). Completed studies. Retrieved September 22, 2007 from http://www.iea.nl/completed_studies.html

Jacob, B. A., & Lefgren, L. (2004). The impact of teacher training on student achievement: Quasi experiment evidence from school reform efforts in Chicago. *The Journal of Human Resources, 39* (1), 50-79.

Johnson, C. C., Kahle, J. B., & Fargo, J. D. (2007). A study of the effect of sustained, whole-school professional development on student achievement in science. *Journal of reseach in science teaching, 44* (6), 775-786.

Keith, T. Z., & Cool, V. A. (1992). Testing models of school learning: Effects of quality of instruction, motivation, academic coursework, and homework on academic achievement. *School Psychology Quarterly, 7,* 207–226. [Could not retrieve full text online] Educational Outcomes of Tutoring: A Meta-Analysis of Findings

Kennedy, M. M. (1998). *Form and substance in in-service teacher education.* Arlington, VA: National Science Foundation.

Kennedy, M. M. (1998). *Form and substance in in-service teacher education (Re- search Monograph No. 13).* Arlington, VA: National Science Foundation.

Kiecolt, K. J., Nathan, L. E. (1985). *Secondary analysis of survey data*. Beverly Hills: Sage.

Koller, Baumert & Schnabel (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education, 32* (5), 448-470.

Kreft, I.G.G. (1996). Are multilevel techniques necessary? An overview, including simulation studies. Unpublished manuscript, California State University, Los Angeles, CA.

Kreft, I., & Leeuw, J. D. (2004). *Introducing Multilevel Modeling*. London: Sage Publications.

Kromrey, J.D. & Hines, C.V. (1994). Nonrandomly missing data in multiple regression: an empirical comparison of common missing-data treatments. *Educational and Psychological Measurement,* 54, 573-593.

Luke, A. D. (2004). *Multilevel Modeling*. Series: Quantitative Applications in the Social Sciences. Sage Publications

Lee, J. (2007). Two worlds of private tutoring: the prevalence and causes of after-school mathematics tutoring in Korea and the United States. *Teachers College Record, 109,* 1207-1234.

Leung, F. K. S. (2002). Behind the high achievement of East Asian students. *Educational Research and Evaluation, 8,* 87–108.

Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis,* 15(2), 129-1 51.

Loucks-Horsley, S., Hewson, P. W., Love, N., & Stiles, K. E. (1998). *Designing professional development for teachers of science and mathematics.* Thousand Oaks, CA: Corwin Press.

Luyten, H., Visscher, A., & Witziers, B. (2005). School effectiveness research: From a review of the criticism to recommendations for further development. *School Effectiveness and School Improvement, 16* (3), 249-279.

Ma, X., & Kishor, N. (1997). Assessing the Relationship Between Attitude Toward Mathematics and Achievement in Mathematics: A Meta-Analysis. *Journal for Research in Mathematics Education, 28* (1), 26-47.

Marsh, H. W., Hau, K., & Kong, C. (2002). Multilevel causal ordering of academic self-concept and achievement: Influence of language of instruction (English compared with Chinese) for Hong Kong students. *American Educational Research Journal, 39,* 727–763.

Martin, M. O. (2005). *TIMSS 2003 user guide for the international database.* Chestnut Hill,MA: TIMSS & PIRLS International Study Center.

Moriarty, H. J., Deatrick, J. A., Mahon, M. M., Feetham, S. L., Carroll , R. M., Shepard, M. P., & Orsi, A. J. (1999). Issues to consider when choosing and using large national databases for research of families. *Western Journal of Nursing Research, 21,* 143-153.

Mosteller, F., & Moynihan, D. P. (Eds.) (1972). *On equality of educational opportunity.* New York: Vintage Books.

Muijs, D., & Reynolds, D. (2003). The effectiveness of the use of learning support assistants in improving the mathematics achievement of low achieving pupils in primary school. *Educational Research, 45* (3), 219-230.

Mullis, S. R. (2001). Dealing with Missing Data. Outcome Oriented: The Online Newsletter of the Center for Outcome Measurement in Brain Injury (COMBI). Retrieved July 14, 2006 from www.tbims.org/combi.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Muthen, B., Huang, L., Jo, B., Khoo, S., Goff, G. N., Novak, J. R., & Shih, J. C. (1995). Opportunity-to-learn effects on achievement: Analytical aspects. *Educational Evaluation and Policy Analysis*, 17 (3), 371-403.

National Center for Education Statistics. (2007). Surveys and programs. Retrieved on June 15, 2007 from http://nces.ed.gov/surveys/international/IntlIndicators/index.asp.

National Commission on Educational Excellence. (1983). A national at risk: The imperative for educational reform. Washington, D.C.: United States Department of Education.

No Child Left Behind. (2001). No Child Left Behind Act of 2001. Pub. L. 107-110. Washington D.C: U.S. Department of Education.

Nye, B., Hedges, L. V., & Konstantopoulos, S. (2001). The long-term effects of small classes in early grades: Lasting benefits in mathematics achievement at grade 9. *The Journal of Experimental Education, 2001, 69* (3), 245-257.

Organization for Economic Co-operation and Development. (2001). *Knowledge and skills for life: First results from the OECD programme for international student assessment*. Paris, France: OECD.

Organization for Economic Co-operation and Development. (2007). The program for international student assessment. Retrieved September 20, 2007 from http://www.pisa.oecd.org.

O'Leary, M. (2002). Stability of country rankings across item formats in the third international mathematics and science study. *Educational Measurement: Issues and Practice, 21* (4), 27-38.

Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology, 24,* 124–139.

Papanastasiou, C. (2000). Effects of attitudes and beliefs on mathematics achievement. *Studies in Educational Evaluation, 26,* 27-42.

Papanastasiou, C. (2002). School, teaching and family influence on student attitudes toward science: Based on TIMSS data Cyprus. *Studies in Educational Evaluation, 28,* 71-86.

Parsad, B., Lewis, L., Farris, E., Greene, B., (2000). *Teacher preparation and professional development: 2000.* U.S. Department of Education: Office of Educational Research and Improvement.

Peterson, P. L., & Fennema, E. (1985). Effective teaching, student engagement in classroom activities, and sex related differences in learning mathematics. *American Educational Research Journal, 22* (3), 309-335.

Phan, H. T., & Kromrey, J. D. (2007). Missing data in large-scale assessments: A confirmatory factor analysis of the TIMSS 2003 eighth-grade mathematics scores. A research paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Pianta, R. C., Belsky, J., Houts, R., & Morrison, F. (2007). Opportunities to learn in America's elementary classrooms. *Science, 315* (5820) 1795-1796.

Pong, S., & Pallas, A. (2001). Class size and eighth-grade math achievement in the United States and Abroad. *Educational Evaluation and Policy Analysis, 23* (3), 251-273.

Rao, N., Moely, B. E., & Sachs, J. (2000). Motivational beliefs, study strategies, and mathematics attainment in high- and low-achieving Chinese secondary school students. *Contemporary Educational Psychology, 25,* 287–316.

Raudenbush, S. W. &  Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd Ed.) Thousand Oaks: Sage.

Raudenbush, S., Bryk, A., Cheong, Y.F., & Congdon, R. (2004). *HLM6 hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International, Inc.

Raynolds, A. J. (1991). The middle schooling process: Influences on science and mathematics achievement from the longitudinal study of American youth. *Adolescence, 26,* 132-157.

Reys, B. J., Reys, R. E., Barnes, D., Beem, J., & Papick, I. (1997). Collaborative curriculum investigation as a vehicle for teacher enhancement and mathematics curriculum reform. *School Science and Mathematics, 87* (5), 253-259.

Riddell, A. R. (1997). Assessing designs for school effectiveness research and school improvement in developing countries. *Comparative Education Review*, *41* (2), 178-204.

Rodriguez, M. C. (2004). The Role of classroom assessment in student performance on TIMSS. *Applied Measurement in Education, 17* (1), 1-24.

Rosenberg, A.L., Greenfield, M.V.H., & Dimick, J.B. (2006). *Secondary data analysis: Using existing data to answer clinical questions*. In J.T. Wei (Ed.) Clinical Research Methods for the Surgeons. Totowa, NJ: Humana Press.

Ross, J. A., Bruce, C., & Hogaboam-Gray, A. (2006). The impact of a professional development program on student achievement in grade 6 mathematics. *Journal of Mathematics Teacher Education, 9,* 551-577.

Ross, J. G., Saavedra, P. J., Shur, G. H., Winters, F., & Felner, R. D. (1992). The effectiveness of an after-school program for primary grade latchkey students on precursors of substance abuse. *Journal of Community Psychology,* 22-38.

Roth, P. (1994). Missing data:A conceptual overview for applied psychologists. *Personnel Psychology,* 47, 537-560.

SAS Institute Inc. (2005). *SAS/STAT user's guide, version 9.13.* Cary, NC: SAS Institute Inc.

SAS/STAT Software Enhancement. (2006). SAS/STAT Software Enhancement. Retrieved on July 14, 2006 from http://support.sas.com/rnd/app/da/new/dami.html.

Schmidt, W. H., & McKnight, C. C. (1995). Surveying Educational Opportunity in Mathematics and Science: An International Perspective. *Educational Evaluation and Policy Analysis*, 17 (3), 337-353.

Schmidt, W., McKnight, C., Valverde, G., Houang, R., & Willey D. (1997). *Many visions, many aims: A cross-national investigation of curricular intentions in school mathematics*. Norwell, MA: Kluwer Academic.

Shen, C. (2002). Revisiting the relationship between students' achievement and their self perceptions: A cross-national analysis based on TIMSS 1999 data. *Assessment in Education, 9* (2), 161–184.

Shen, C., & Pedulla, J. J. (2000). The relationship between students' achievement and their self perceptions of competence and rigour of mathematics and science: A cross-national analysis. *Assessment in Education, 7*, 237–253.

Shulman, L. S., & Shulman, J, H. (2004). How and what teachers learn: a shifting perspective. *Journal of Curriculum Studies, 36* (2), 257–271.

Singh, K., Grandville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement, *The Journal of Educational Research, 95* (6), 323-332.

Slavin, R. E. (1989). Class size and student achievement: Small effects of small classes. *Educational Psychologist, 24* (1), 99-110.

Smith, D. C., & Neale, D. C. (1991). The construction of subject-matter knowledge in primary science teaching. In J. Brophy (Ed.), *Advances in research on teaching: Vol. 2. Teachers' knowledge of subject matter as it relates to their teaching practice* (18 –243). Greenwich, CT: JAI Press.

Staub, F.C., & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology*, *94*, 344-355.

Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS videotape classroom study: Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States.* U.S. Department of Education. National Center for Education Statistics NCES 1999-074. Washington, DC: U.S. Government Printing Office.

Stigler, J. W., Gallimore, R., & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS video studies. *Educational Psychologist, 35,* 87-100.

Stigler, J. W., Lee, S. Y., & Stevenson, H. W. (1987). Mathematics classroom in Japan, Taiwan, and the United States. *Child Development, 58,* 1272-1285.

Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education, 17,* 213-226.

Suter, L. E. (2000). Is student achievement immutable? Evidence from international studies on schooling and student achievement. *Review of Educational Research, 70* (4), 529-545.

The World Bank (2003). *2003 World Development Indicators*. Washington, DC: The World Bank.

The World Bank. (2007). World Bank list of economies (July 2007). Retrieved on November 30, 2007 from http://web.worldbank.org/wbsite/external/datastatistics/0,,contentmdk:20420458~menupk:64133156~pagepk:64133150~pipk:64133175 ~thesitepk:239419,00.html

TIMSS. (1995). The trends in international mathematics and science study (TIMSS 1995). Retrieved on June 15, 2007 from http://isc.bc.edu/timss1995.html.

TIMSS. (1999). Mathematics Benchmarking Report: TIMSS 1999. Retrieved on June 15, 2007 from http://isc.bc.edu/timss1999b/pdf/TB99_Math_Intro.pdf

TIMSS. (2003). TIMSS 2003 Technical Report. Retrieved on June 15, 2007 from http://timss.bc.edu/timss2003i/technicalD.html

TIMSS. (2007). The trends in international mathematics and science study (TIMSS 2007). Retrieved on June 15, 2007 from http://timss.bc.edu/TIMSS2007/index.html

Trautwein, U., Koller, O., Schmitz, B., & Baumert, J. (2002). Do homework assignments enhance achievement? A multilevel analysis in 7th grade mathematics. *Contemporary Educational Psychology, 27*, 26-50.

Trautwein, U., & Koller, O. (2003). The relationship between homework and achievement - still much of a mystery. *Educational Psychology Review*, Vol. 15, No. 2, 115-145.

Trautwein, U. (2007). The homework-achievement relation reconsidered: Differentiating homework time, homework frequency, and homework effort. *Learning and Instruction 17*, 372-388.

Tschannen-Moran, M. & Hoy, A. W (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17,* 783-805.

Walberg, H.J., & Paschal, R.A. (1995). *Homework*. In L.W. Anderson (Ed.), International encyclopedia of teaching and teacher education (pp. 268–271). Oxford: Elsevier.

Wang, J. (2001). TIMSS primary and middle school data: Some technical concerns. *Educational Researcher, 30* (6), 17-21.

Watkins, D. A., & Biggs, J. B. (2001). *Teaching the Chinese learners: Psychological and pedagogical perspective.* Hong Kong: Comparative Education Research Centre, the University of Hong Kong and Melbourne, Australia: The Australian Council for Educational Research.

Werf, G. V. D., Creemers, B., Jong, R. D., & Klaver, E. (2000). Evaluation of school improvement through an educational effectiveness model: The case of Indonesia's PEQIP Project. *Comparative Education Review, 44* (3), 329-355.

Westbury, I. (1992). Comparing American and Japanese achievement: Is the United States really a low achiever? Educational Researcher, 21(4), 18-24.

Wester, A., & Henriksson, W. (2000). The interaction between item format and gender differences in mathematics performance based on TIMSS data. *Studies in Educational Evaluation*, *26,* 79-90.

Wiggins, R. A., & Follo, E. J. (1999). Development of knowledge, attitudes, and commitment to teach diverse student populations. *Journal of Teacher Education, 50* (2), 94-105.

Wiley, D. E., & Yoon, B. (1995). Teacher reports on opportunity to learn: Analyses of the 1993 california learning assessment system (CLAS). *Educational Evaluation and Policy Analysis, 17* (3), 355-370.

Wobmann, L. (2003). School resources, educational institutions and student performance: The international evidence. Oxford Bulletin of Economics and Statistics. 65(2), 0305-9049.

Wobmann, L., & West, M. (2006). Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS. *European Economic Review, 50,* 695-736.

Wright, S.P., Horn, S.P. & Sanders, L. (1997). Teacher and classroom context effects on students achievement: *Implications for Teacher Evaluation, 11*, 57-67.

Yair, G. (2000). Not just about time: Instructional practices and productive time in school. *Educational Administration Quarterly, 36* (4), 485-512.

APPENDICES

# Appendix A: List of Countries

Table A-1.
*List of countries participating in TIMSS 2003 eighth-grade assessment by country status*

| Developing Countries | | | Developed Countries | | |
|---|---|---|---|---|---|
| Name | No. schools | No. Students | Name | No. schools | No. Students |
| Armenia | 149 | 5,699 | Australia | 207 | 4,791 |
| Bulgaria | 164 | 4,117 | Bahrain | 67 | 4,199 |
| Botswana | 146 | 5,150 | Belgium | 144 | 4,970 |
| Chile | 195 | 6,377 | Canada | 361 | 8,628 |
| Egypt | 217 | 7,094 | Chinese Taipei | 150 | 5,379 |
| Estonia | 151 | 4,040 | Cyprus | 59 | 4,009 |
| Hungary | 155 | 3,302 | United Kingdom | 215 | 6,346 |
| Indonesia | 150 | 5,762 | Hong Kong SAR | 125 | 4,972 |
| Iran | 181 | 4,942 | Israel | 146 | 4,318 |
| Jordan | 140 | 4,489 | Italy | 171 | 4,278 |
| Latvia | 140 | 3,629 | Japan | 146 | 4,856 |
| Lebanon | 152 | 3,814 | Korea | 149 | 5,309 |
| Lithuania | 143 | 4,572 | Netherlands | 130 | 3,036 |
| Malaysia | 150 | 5,314 | New Zealand | 169 | 3,800 |
| Macedonia | 147 | 3,893 | Norway | 138 | 4,133 |
| Morocco | 131 | 2,873 | Singapore | 164 | 6,018 |
| Moldova | 149 | 4,033 | Slovenia | 174 | 3,578 |
| Philippines | 137 | 6,917 | Sweden | 159 | 4,255 |
| Romania | 148 | 4,103 | United States | 286 | 11,100 |
| Russian Federation | 214 | 4,667 | Spain | 115 | 2,514 |
| Saudi Arabia | 155 | 4,295 | | | |
| Serbia | 149 | 4,295 | | | |
| South Africa | 255 | 8,840 | | | |
| Syrian Arab Republic | Complete data not available | | | | |
| Slovak Republic | 179 | 4,215 | | | |
| Tunisia | 150 | 4,931 | | | |
| Palestine | 145 | 5,357 | | | |

Note: The classification of country status was based on the World Bank's (2003) World Development Indicators (The World Bank, 2003). According to the World Bank's (2007) list of economies, developing countries refer to countries with low-income and middle-income economies; whereas developed countries refer to countries with high-income economies. The use of the terms is convenient; it is not intended to imply that all economies in the group are experiencing similar development or that developed economies have reached a preferred or final stage of development (The World Bank, 2007).

## Appendix B: Items Used to Create Composite Variable Opportunity to Learn

*Items used to create composite variable opportunity to learn*

The following list includes the main topics addressed by the TIMSS mathematics test. Choose the response that best describes when students in the TIMSS class have been taught each topic. If a topic was taught half this year and half before this year, please choose "Mostly taught this year."

1 = mostly taught before this year

2 = mostly taught this year

3 = not yet taught or just introduced

A. Number

a) Whole numbers including place value, factorization, and the four operations

b) Computations, estimations, or approximations involving whole numbers

c) Common fractions including equivalent fractions, and ordering of fractions

d) Decimal fractions including place value, ordering, rounding, and converting to common fractions (and vice versa)

e) Representing decimals and fractions using words, numbers, or models (including number lines)

f) Computations with fractions

g) Computations with decimals

h) Integers including words, numbers, or models (including number lines), ordering integers, addition, subtraction, multiplication, and division with integers

i) Ratios (equivalence, division of a quantity by a given ratio)

j) Conversion of percents to fractions or decimals, and vice versa

B. Algebra

a) Numeric, algebraic, and geometric patterns or sequences (extension, missing terms, generalization of patterns)

b) Sums, products, and powers of expressions containing variables

c) Simple linear equations and inequalities, and simultaneous (two variables) equations

d) Equivalent representations of functions as ordered pairs, tables, graphs, words, or equations

e) Proportional, linear, and nonlinear relationships (travel graphs and simple piecewise functions included)

f) Attributes of a graph such as intercepts on axes, and intervals where the function increases, decreases, or is constant

C. Measurement

a) Standard units for measures of length, area, volume, perimeter, circumference, time, speed, density, angle, mass/weight

b) Relationships among units for conversions within systems of units, and for rates

c) Use standard tools to measure length, weight, time, speed, angle, and temperature

d) Estimations of length, circumference, area, volume, weight, time, angle, and speed in problem situations (e.g., circumference of a wheel, speed of a runner)

e) Computations with measurements in problem situations (e.g., add measures, find average speed on a trip, find population density)

f) Measurement formulas for perimeter of a rectangle, circumference of a circle, areas of plane figures (including circles), surface area and volume of rectangular solids, and rates

g) Measures of irregular or compound areas (e.g., by using grids or dissecting and rearranging pieces)

h) Precision of measurements (e.g., upper and lower bounds of a length reported as 8 centimeters to the nearest centimeter)

D. Geometry

a) Angles - acute, right, straight, obtuse, reflex, complementary, and supplementary

b) Relationships for angles at a point, angles on a line, vertically opposite angles, angles associated with a transversal cutting parallel lines, and perpendicularity

c) Properties of angle bisectors and perpendicular bisectors of lines

d) Properties of geometric shapes: triangles and quadrilaterals

e) Properties of other polygons (regular pentagon, hexagon, octagon, decagon)

f) Construct or draw triangles and rectangles of given dimensions

g) Pythagorean theorem (not proof) to find length of a side

h) Congruent figures (triangles, quadrilaterals) and their corresponding measures

i) Similar triangles and recall their properties

j) Cartesian plane - ordered pairs, equations, intercepts, intersections, and gradient

k) Relationships between two-dimensional and three-dimensional shapes

l) Line and rotational symmetry for two-dimensional shapes

m) Translation, reflection, rotation, and enlargement

E. Data

a) Organizing a set of data by one or more characteristics using a tally chart, table, or graph

b) Sources of error in collecting and organizing data (e.g., bias, inappropriate grouping)

c) Data collection methods (e.g., survey, experiment, questionnaire)

d) Drawing and interpreting graphs, tables, pictographs, bar graphs, pie charts, and line graphs

e) Characteristics of data sets including mean, median, range, and shape of distribution (in general terms)

f) Interpreting data sets (e.g., draw conclusions, make predictions, and estimate values between and beyond given data points)

g) Evaluating interpretations of data with respect to correctness and completeness of interpretation

h) Simple probability including using data from experiments to estimate probabilities for favorable outcomes

## Appendix C: Items Used to Create Composite Variable Ready to Teach Math Topics

*Items used to create composite variable ready to teach math topics*

Considering your training and experience in both mathematics content and instruction, how ready do you feel you are to teach each topic at eighth grade? ( 3-point scale: 1 = very ready, 2 = ready, 3 = not ready)

A. Number

1) Representing decimals and fractions using words, numbers, or models (including number lines)

2) Integers including words, numbers, or models (including number lines); ordering integers; and addition, subtraction, multiplication, and division with integers

B. Algebra

1) Numeric, algebraic, and geometric patterns or sequences (extension, missing terms, generalization of patterns)

2) Simple linear equations and inequalities, and simultaneous (two variables) equations

3) Equivalent representations of functions as ordered pairs, tables, graphs, words, or equations

4) Attributes of a graph such as intercepts on axes, and intervals where the function increases, decreases, or is constant

C. Measurement

1) Estimations of length, circumference, area, volume, weight, time, angle, and speed in problem situations (e.g., circumference of a wheel, speed of a runner)

2) Computations with measurements in problem situations (e.g., add measures, find average speed on a trip, find population density)

3) Measures of irregular or compound areas (e.g., by using grids or dissecting and rearranging pieces)

4) Precision of measurements (e.g., upper and lower bounds of a length reported as 8 centimeters to the nearest centimeter)

D. Geometry

1) Pythagorean theorem (not proof) to find length of a side

2) Congruent figures (triangles, quadrilaterals) and their corresponding measures

3) Cartesian plane - ordered pairs, equations, intercepts, intersections, and gradient

4) Translation, reflection, rotation, and enlargement

E. Data

1) Sources of error in collecting and organizing data (e.g., bias, inappropriate grouping)

2) Data collection methods (e.g., survey, experiment, questionnaire)

3) Characteristics of data sets including mean, median, range, and shape of distribution (in general terms)

4) Simple probability including using data from experiments to estimate probabilities for favorable outcomes

# Appendix D: Reliabilities of Composite Variables

Table A-2.
*Factor pattern coefficients of items used to create composite variables*

| Composite variable | Item description | Factor pattern coefficients | Cronbach's alpha |
|---|---|---|---|
| Student self-confidence (TIMSS derived variable) | I learn things quickly in math | 0.65 | 0.73 |
| | I usually do well in math | 0.65 | |
| | Math is more difficult for me than for many of my classmates | 0.56 | |
| | Math is not one of my strengths | 0.64 | |
| Student valuing math (TIMSS derived variable) | I need math to learn other school subjects | 0.69 | 0.79 |
| | I need to do well in math t get the job I want | 0.65 | |
| | I would like a job that involved using math | 0.64 | |
| | I need to do well in math to get into the university of my choice | 0.62 | |
| | I would like to take more math in school | 0.59 | |
| | I think learning math will help me in my daily life | 0.46 | |
| | I enjoy learning math | 0.50 | |
| Home resources for learning | Desk | 0.44 | 0.44 |
| | Calculator | 0.39 | |
| | Computer | 0.42 | |
| Opportunity to learn number (TIMSS derived variable) | Whole numbers including place value, factorization, and the four operations | 0.66 | 0.91 |
| | Computations, estimations, or approximations involving whole numbers | 0.72 | |
| | Common fractions including equivalent fractions, and ordering of fractions | 0.74 | |
| | Decimal fractions including place value, ordering, rounding, and converting to common fractions (and vice versa) | 0.74 | |
| | Representing decimals and fractions using words, numbers, or models (including number lines) | 0.74 | |
| | Computations with fractions | 0.81 | |
| | Computations with decimals | 0.80 | |
| | Integers including words, numbers, or models (including number lines), ordering integers, addition, subtraction, multiplication, and division with integers | 0.60 | |
| | Ratios (equivalence, division of a quantity by a given ratio) | 0.64 | |
| | Conversion of percents to fractions or decimals, and vice versa | 0.72 | |

Table A-2.
*Factor pattern coefficients of items used to create composite variables*

| Composite variable | Item description | Factor pattern coefficients | Cronbach's alpha |
|---|---|---|---|
| Opportunity to learn algebra (TIMSS derived variable) | Numeric, algebraic, and geometric patterns or sequences (extension, missing terms, generalization of patterns) | 0.36 | 0.74 |
| | Sums, products, and powers of expressions containing variables | 0.46 | |
| | Simple linear equations and inequalities, and simultaneous (two variables) equations | 0.61 | |
| | Equivalent representations of functions as ordered pairs, tables, graphs, words, or equations | 0.67 | |
| | Proportional, linear, and nonlinear relationships (travel graphs and simple piecewise functions included) | 0.63 | |
| | Attributes of a graph such as intercepts on axes, and intervals where the function increases, decreases, or is constant | 0.68 | |
| Opportunity to learn measurement (TIMSS derived variable) | Standard units for measures of length, area, volume, perimeter, circumference, time, speed, density, angle, mass/weight | 0.68 | 0.85 |
| | Relationships among units for conversions within systems of units, and for rates | 0.66 | |
| | Use standard tools to measure length, weight, time, speed, angle, and temperature | 0.67 | |
| | Estimations of length, circumference, area, volume, weight, time, angle, and speed in problem situations (e.g., circumference of a wheel, speed of a runner) | 0.67 | |
| | Computations with measurements in problem situations (e.g., add measures, find average speed on a trip, find population density) | 0.66 | |
| | Measurement formulas for perimeter of a rectangle, circumference of a circle, areas of plane figures (including circles), surface area and volume of rectangular solids, and rates | 0.65 | |
| | Measures of irregular or compound areas (e.g., by using grids or dissecting and rearranging pieces) | 0.62 | |
| | Precision of measurements (e.g., upper and lower bounds of a length reported as 8 centimeters to the nearest centimeter) | 0.57 | |

## Appendix D: (Continued)

Table A-2.
*Factor pattern coefficients of items used to create composite variables*

| Composite variable | Item description | Factor pattern coefficients | Cronbach's alpha |
|---|---|---|---|
| Opportunity to learn geometry (TIMSS derived variable) | Angles - acute, right, straight, obtuse, reflex, complementary, and supplementary | 0.67 | 0.88 |
| | Relationships for angles at a point, angles on a line, vertically opposite angles, angles associated with a transversal cutting parallel lines, and perpendicularity | 0.64 | |
| | Properties of angle bisectors and perpendicular bisectors of lines | 0.67 | |
| | Properties of geometric shapes: triangles and quadrilaterals | 0.74 | |
| | Properties of other polygons (regular pentagon, hexagon, octagon, decagon) | 0.69 | |
| | Construct or draw triangles and rectangles of given dimensions | 0.65 | |
| | Pythagorean theorem (not proof) to find length of a side | 0.48 | |
| | Congruent figures (triangles, quadrilaterals) and their corresponding measures | 0.75 | |
| | Similar triangles and recall their properties | 0.59 | |
| | Cartesian plane - ordered pairs, equations, intercepts, intersections, and gradient | 0.51 | |
| | Relationships between two-dimensional and three-dimensional shapes | 0.48 | |
| | Line and rotational symmetry for two-dimensional shapes | 0.57 | |
| | Translation, reflection, rotation, and enlargement | 0.63 | |

Table A-2.
*Factor pattern coefficients of items used to create composite variables*

| Composite variable | Item description | Factor pattern coefficients | Cronbach's alpha |
|---|---|---|---|
| Opportunity to learn data (TIMSS derived variable) | Organizing a set of data by one or more characteristics using a tally chart, table, or graph | 0.64 | 0.85 |
| | Sources of error in collecting and organizing data (e.g., bias, inappropriate grouping) | 0.64 | |
| | Data collection methods (e.g., survey, experiment, questionnaire) | 0.67 | |
| | Drawing and interpreting graphs, tables, pictographs, bar graphs, pie charts, and line graphs | 0.63 | |
| | Characteristics of data sets including mean, median, range, and shape of distribution (in general terms) | 0.63 | |
| | Interpreting data sets (e.g., draw conclusions, make predictions, and estimate values between and beyond given data points) | 0.73 | |
| | Evaluating interpretations of data with respect to correctness and completeness of interpretation | 0.70 | |
| | Simple probability including using data from experiments to estimate probabilities for favorable outcomes | 0.59 | |
| Instructional practice-related activities in math lessons | We relate what we are learning in mathematics to our daily life | 0.51 | 0.55 |
| | We decide on our own procedures for solving complex problems | 0.47 | |
| | We work together in small groups | 0.39 | |
| | We explain our answers | 0.40 | |
| | We listen to the teacher give a lecture-style presentation | 0.36 | |
| Content-related activities in math lesson | We practice adding, subtracting, multiplying, and dividing without using a calculator | 0.42 | 0.60 |
| | We work on fractions and decimals | 0.55 | |
| | We interpret data in tables, charts, or graphs | 0.53 | |
| | We write equations and functions to represent relationships | 0.50 | |
| Ready to teach number | Representing decimals and fractions using words, numbers, or models | 0.66 | 0.71 |
| | Integers including words, numbers, or models (including number lines); ordering integers; and addition, subtraction, multiplication, and division with integers | 0.66 | |

Table A-2.
*Factor pattern coefficients of items used to create composite variables*

| Composite variable | Item description | Factor pattern coefficients | Cronbach's alpha |
|---|---|---|---|
| Ready to teach algebra | Numeric, algebraic, and geometric patterns or sequences (extension, missing terms, generalization of patterns) | 0.57 | 0.81 |
| | Simple linear equations and inequalities, and simultaneous (two variables) equations | 0.73 | |
| | Equivalent representations of functions as ordered pairs, tables, graphs, words, or equations | 0.79 | |
| | Attributes of a graph such as intercepts on axes, and intervals where the function increases, decreases, or is constant | 0.73 | |
| Ready to teach measurement | Estimations of length, circumference, area, volume, weight, time, angle, and speed in problem situations (e.g., circumference of a wheel, speed of a runner) | 0.79 | 0.86 |
| | Computations with measurements in problem situations (e.g., add measures, find average speed on a trip, find population density) | 0.79 | |
| | Measures of irregular or compound areas (e.g., by using grids or dissecting and rearranging pieces) | 0.77 | |
| | Precision of measurements (e.g., upper and lower bounds of a length reported as 8 centimeters to the nearest centimeter) | 0.73 | |
| Ready to teach geometry | Pythagorean theorem (not proof) to find length of a side | 0.82 | 0.83 |
| | Congruent figures (triangles, quadrilaterals) and their corresponding measures | 0.83 | |
| | Cartesian plane - ordered pairs, equations, intercepts, intersections, and gradient | 0.72 | |
| | Translation, reflection, rotation, and enlargement | 0.66 | |
| Ready to teach data | Sources of error in collecting and organizing data (e.g., bias, inappropriate grouping) | 0.81 | 0.83 |
| | Data collection methods (e.g., survey, experiment, questionnaire) | 0.80 | |
| | Characteristics of data sets including mean, median, range, and shape of distribution (in general terms) | 0.69 | |
| | Simple probability including using data from experiments to estimate probabilities for favorable outcomes | 0.66 | |

Table A-2.
*Factor pattern coefficients of items used to create composite variables*

| Composite variable | Item description | Factor pattern coefficients | Cronbach's alpha |
|---|---|---|---|
| Math-related professional development | Math content | 0.74 | 0.78 |
| | Math pedagogy/instruction | 0.69 | |
| | Math curriculum | 0.69 | |
| | Math assessment | 0.54 | |
| | Problem solving/critical thinking | 0.50 | |
| School resources for mathematics instruction (TIMSS derived variable) | Computers for mathematics instruction | 0.82 | 0.92 |
| | Computer software for mathematics instruction | 0.82 | |
| | Audio-visual resources for mathematics instruction | 0.84 | |
| | Library materials relevant to mathematics instruction | 0.82 | |
| | Calculators for mathematics instruction | 0.80 | |
| | Budget for supplies (e.g., paper, pencils) | 0.68 | |
| | School buildings and grounds | 0.67 | |
| | Heating/cooling and lighting systems | 0.68 | |
| | Instructional materials (e.g., textbook) | 0.66 | |
| | Instructional space (e.g., classrooms) | 0.61 | |
| Teacher's perception of math instructional limitations due to student factors (TIMSS derived variable) | Low morale among students | 0.80 | 0.81 |
| | Uninterested students | 0.76 | |
| | Disruptive students | 0.69 | |
| | Students with special needs, (e.g., hearing, vision, speech impairment, physical disabilities, mental or emotional/psychological impairment) | 0.54 | |
| | Students who come from a wide range of backgrounds (e.g., economic, language) | 0.56 | |
| | Students with different academic abilities | 0.52 | |

# Appendix E: Weighted Correlation of Level-1 Variables for USA

Table A-3.
*Weighted Correlation of Level-1 Variables for USA (N = 4,414)*

| Variable | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Gender | 1.00 | | | | | |
| 2. Self-confidence in learning math | -.15 | 1.00 | | | | |
| 3. Valuing of math | -.06 | .39 | 1.00 | | | |
| 4. Time on math homework | -.06 | .06 | .01 | 1.00 | | |
| 5. Extra math lessons | -.04 | -.13 | .01 | -.01 | 1.00 | |
| 6. Home resources for learning math | .05 | .10 | .12 | -.04 | -.04 | 1.00 |

Appendix F: Unweighted Correlation of Level-2 Variables for USA

Table A-4.
*Unweighted Correlation of Level-2 Variables for USA (N = 153)*

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Opportunity_ number | 1.00 | | | | | | | | | | | | | | | | | | |
| 2. Opportunity_ algebra | .19 | 1.00 | | | | | | | | | | | | | | | | | |
| 3. Opportunity_ measurement | .20 | .45 | 1.00 | | | | | | | | | | | | | | | | |
| 4. Opportunity_ geometry | .09 | .25 | .41 | 1.00 | | | | | | | | | | | | | | | |
| 5. Opportunity_data | .00 | .35 | .28 | .37 | 1.00 | | | | | | | | | | | | | | |
| 6. Homework assignment | -.01 | -.05 | .13 | -.02 | .03 | 1.00 | | | | | | | | | | | | | |
| 7. Instruction_ activities | .05 | .08 | .08 | -.04 | .06 | -.07 | 1.00 | | | | | | | | | | | | |
| 8. Content_ activities | .09 | .11 | .16 | -.10 | -.07 | .18 | .31 | 1.00 | | | | | | | | | | | |
| 9. Preparation to teach | .18 | .03 | .17 | .17 | .05 | -.03 | .02 | .00 | 1.00 | | | | | | | | | | |
| 10. Ready_number | .37 | .15 | .15 | .03 | .02 | .13 | .09 | .14 | .19 | 1.00 | | | | | | | | | |
| 11. Ready_algebra | .18 | .27 | .24 | .06 | .03 | .15 | -.04 | .21 | .26 | .54 | 1.00 | | | | | | | | |
| 12. Ready_ measurement | .17 | .23 | .38 | .19 | .16 | .16 | -.01 | .23 | .21 | .59 | .70 | 1.00 | | | | | | | |
| 13. Ready_geometry | .19 | .07 | .30 | .10 | .04 | .23 | -.01 | .05 | .34 | .60 | .47 | .61 | 1.00 | | | | | | |
| 14. Ready_data | .25 | .08 | .32 | .02 | .10 | .23 | .09 | .15 | .23 | .55 | .52 | .51 | .53 | 1.00 | | | | | |
| 15. Professional development | .06 | .15 | .13 | .14 | .23 | .00 | .10 | .07 | .10 | .02 | .14 | .11 | .03 | .03 | 1.00 | | | | |
| 16. Class size | .03 | .13 | .11 | .06 | .14 | -.05 | -.01 | .07 | .13 | .02 | -.01 | -.08 | -.12 | .08 | .11 | 1.00 | | | |
| 17. School resources | .06 | .04 | .11 | .06 | .07 | .08 | .08 | .09 | .18 | .00 | .18 | .18 | .11 | .06 | -.09 | .04 | 1.00 | | |
| 18. Teacher perception_limitation | -.02 | -.19 | -.18 | -.06 | -.15 | -.31 | -.05 | -.03 | -.12 | .15 | .07 | .10 | -.04 | .00 | .02 | .07 | -.15 | 1.00 | |
| 19. Math hours per year | .11 | .11 | .15 | -.03 | -.04 | -.02 | .00 | .07 | .03 | -.06 | .04 | .02 | -.05 | -.09 | .11 | .10 | -.02 | -.07 | 1.00 |

## Appendix G: Weighted Correlation of Level-1 Variables for Canada

Table A-5.
*Weighted Correlation of Level-1 Variables for Canada (N = 6,248)*

| Variable | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Gender | 1.00 | | | | | |
| 2. Self-confidence in learning math | -.11 | 1.00 | | | | |
| 3. Valuing of math | -.04 | .37 | 1.00 | | | |
| 4. Time on math homework | -.02 | .04 | -.03 | 1.00 | | |
| 5. Extra math lessons | -.02 | -.24 | -.01 | -.06 | 1.00 | |
| 6. Home resources for learning math | .03 | .09 | .14 | -.04 | -.01 | 1.00 |

# Appendix H: Unweighted Correlation of Level-2 Variables for Canada

Table A-6.
*Unweighted Correlation of Level-2 Variables for Canada (N = 271)*

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Opportunity_ number | 1.00 | | | | | | | | | | | | | | | | | | |
| 2. Opportunity_ algebra | .14 | 1.00 | | | | | | | | | | | | | | | | | |
| 3. Opportunity_ measurement | -.08 | .20 | 1.00 | | | | | | | | | | | | | | | | |
| 4. Opportunity_ geometry | .03 | .19 | .37 | 1.00 | | | | | | | | | | | | | | | |
| 5. Opportunity_data | -.06 | .18 | .47 | .41 | 1.00 | | | | | | | | | | | | | | |
| 6. Homework assignment | .02 | .02 | .15 | .06 | .08 | 1.00 | | | | | | | | | | | | | |
| 7. Instruction_ activities | .06 | .01 | .09 | .11 | .08 | .02 | 1.00 | | | | | | | | | | | | |
| 8. Content_ activities | -.08 | .11 | .14 | .10 | .19 | .11 | .40 | 1.00 | | | | | | | | | | | |
| 9. Preparation to teach | .21 | .07 | -.27 | -.20 | -.30 | -.07 | -.06 | -.09 | 1.00 | | | | | | | | | | |
| 1. Ready_number | .01 | -.07 | .00 | .00 | -.01 | -.02 | .03 | .06 | .06 | 1.00 | | | | | | | | | |
| 11. Ready_algebra | .03 | .15 | .09 | .04 | .04 | -.09 | .05 | .07 | .17 | .49 | 1.00 | | | | | | | | |
| 12. Ready_ measurement | -.02 | .02 | .12 | .09 | -.03 | .01 | .09 | .12 | .12 | .45 | .53 | 1.00 | | | | | | | |
| 13. Ready_geometry | .04 | -.03 | -.06 | .09 | -.07 | -.13 | .08 | .05 | .16 | .50 | .49 | .50 | 1.00 | | | | | | |
| 14. Ready_data | .15 | .01 | .11 | .09 | .23 | -.01 | .02 | .04 | .02 | .39 | .46 | .41 | .42 | 1.00 | | | | | |
| 15. Professional development | -.12 | .05 | .23 | .09 | .24 | .15 | .01 | .10 | -.13 | .07 | .16 | .15 | .09 | .10 | 1.00 | | | | |
| 16. Class size | .10 | -.06 | -.07 | -.06 | -.06 | .06 | -.04 | .01 | .16 | -.11 | -.06 | .02 | .00 | -.07 | -.15 | 1.00 | | | |
| 17. School resources | .08 | .02 | -.13 | -.01 | -.09 | -.09 | .13 | .03 | .15 | .02 | .01 | .09 | .16 | -.01 | -.06 | .00 | 1.00 | | |
| 18. Teacher perception_limitation | -.09 | .00 | -.04 | -.06 | .09 | .04 | -.08 | -.01 | -.02 | -.11 | -.02 | -.05 | -.09 | .06 | -.05 | .04 | -.04 | 1.00 | |
| 19. Math hours per year | .09 | .10 | .09 | .07 | .14 | .00 | -.04 | -.04 | -.07 | -.03 | -.05 | -.07 | .00 | .04 | .06 | -.12 | -.06 | .02 | 1.00 |

## Appendix I: Weighted Correlation of Level-1 Variables for Egypt

Table A-7.
*Weighted Correlation of Level-1 Variables for Egypt (N = 1,876)*

| Variable | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Gender | 1.00 | | | | | |
| 2. Self-confidence in learning math | -.06 | 1.00 | | | | |
| 3. Valuing of math | .00 | .33 | 1.00 | | | |
| 4. Time on math homework | .02 | -.05 | -.04 | 1.00 | | |
| 5. Extra math lessons | -.02 | -.04 | -.03 | -.01 | 1.00 | |
| 6. Home resources for learning math | 0.01 | 0.13 | 0.07 | 0.02 | -0.03 | 1.00 |

# Appendix K: Unweighted Correlation of Level-2 Variables for Egypt

Table A-8.
*Unweighted Correlation of Level-2 Variables for Egypt (N = 69)*

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Opportunity_ number | 1.00 | | | | | | | | | | | | | | | | | | |
| 2. Opportunity_ algebra | .32 | 1.00 | | | | | | | | | | | | | | | | | |
| 3. Opportunity_ measurement | .20 | .48 | 1.00 | | | | | | | | | | | | | | | | |
| 4. Opportunity_ geometry | .31 | .17 | .39 | 1.00 | | | | | | | | | | | | | | | |
| 5. Opportunity_data | .07 | .25 | .32 | .27 | 1.00 | | | | | | | | | | | | | | |
| 6. Homework assignment | .06 | -.01 | .12 | -.02 | .22 | 1.00 | | | | | | | | | | | | | |
| 7. Instruction_ activities | -.06 | .09 | -.11 | -.05 | -.11 | .04 | 1.00 | | | | | | | | | | | | |
| 8. Content_ activities | .15 | .08 | -.07 | .19 | .12 | .00 | .27 | 1.00 | | | | | | | | | | | |
| 9. Preparation to teach | | | | | | | | | | | | | | | | | | | |
| 1. Ready_number | -.07 | .05 | -.06 | -.13 | .07 | .10 | -.10 | .06 | | 1.00 | | | | | | | | | |
| 11. Ready_algebra | -.08 | .23 | .28 | .02 | .16 | .27 | -.05 | .02 | | .56 | 1.00 | | | | | | | | |
| 12. Ready_ measurement | -.12 | .07 | .05 | -.04 | .24 | .20 | -.12 | -.13 | | .62 | .52 | 1.00 | | | | | | | |
| 13. Ready_geometry | -.09 | -.10 | -.06 | .04 | .19 | .24 | -.18 | .03 | | .21 | .17 | .44 | 1.00 | | | | | | |
| 14. Ready_data | -.06 | .08 | .30 | .00 | .19 | .14 | -.22 | -.10 | | .17 | .31 | .38 | .22 | 1.00 | | | | | |
| 15. Professional development | -.16 | .08 | .02 | .02 | .07 | .06 | -.03 | .11 | | .16 | .23 | .22 | .05 | .12 | 1.00 | | | | |
| 16. Class size | -.02 | -.15 | -.01 | -.20 | -.14 | .04 | -.04 | -.16 | | -.09 | .05 | -.30 | -.20 | -.01 | -.20 | 1.00 | | | |
| 17. School resources | -.23 | -.05 | .05 | -.11 | -.02 | -.07 | .09 | .04 | | .07 | .25 | .23 | -.05 | -.01 | .00 | -.06 | 1.00 | | |
| 18. Teacher perception_limitation | .04 | .07 | .14 | .05 | .00 | .03 | -.12 | -.16 | | -.05 | -.01 | -.11 | -.10 | -.03 | -.10 | .14 | .16 | 1.00 | |
| 19. Math hours per year | -.07 | .04 | -.04 | -.23 | -.06 | .17 | .18 | .00 | | .01 | .15 | .09 | -.07 | -.01 | .06 | .23 | .23 | -.14 | 1.00 |

## Appendix L: Weighted Correlation of Level-1 Variables for South Africa

Table A-9.
*Weighted Correlation of Level-1 Variables for South Africa (N = 1,564)*

| Variable | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Gender | 1.00 | | | | | |
| 2. Self-confidence in learning math | -.07 | 1.00 | | | | |
| 3. Valuing of math | .05 | .18 | 1.00 | | | |
| 4. Time on math homework | .02 | -.02 | .00 | 1.00 | | |
| 5. Extra math lessons | -.03 | -.06 | .03 | -.02 | 1.00 | |
| 6. Home resources for learning math | -.02 | .09 | .03 | -.07 | -.05 | 1.00 |

## Appendix M: Unweighted Correlation of Level-2 Variables for South Africa

Table A-10.

*Unweighted Correlation of Level-2 Variables for South Africa (N = 52)*

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Opportunity_ number | 1.00 | | | | | | | | | | | | | | | | | | |
| 2. Opportunity_ algebra | .29 | 1.00 | | | | | | | | | | | | | | | | | |
| 3. Opportunity_ measurement | .49 | .28 | 1.00 | | | | | | | | | | | | | | | | |
| 4. Opportunity_ geometry | .40 | .63 | .42 | 1.00 | | | | | | | | | | | | | | | |
| 5. Opportunity_data | .31 | .28 | .36 | .45 | 1.00 | | | | | | | | | | | | | | |
| 6. Homework assignment | .12 | .09 | -.02 | .04 | -.05 | 1.00 | | | | | | | | | | | | | |
| 7. Instruction_ activities | -.37 | .01 | -.15 | .06 | -.05 | -.24 | 1.00 | | | | | | | | | | | | |
| 8. Content_ activities | .01 | -.10 | .02 | .06 | .02 | .12 | .40 | 1.00 | | | | | | | | | | | |
| 9. Preparation to teach | .19 | .01 | .15 | .33 | .05 | .14 | -.15 | -.06 | 1.00 | | | | | | | | | | |
| 1. Ready_number | .13 | .09 | .12 | .28 | .13 | .13 | .02 | -.09 | .21 | 1.00 | | | | | | | | | |
| 11. Ready_algebra | .32 | .11 | .11 | .27 | .01 | .09 | .03 | .13 | .13 | .46 | 1.00 | | | | | | | | |
| 12. Ready_ measurement | .43 | .14 | .22 | .25 | .17 | .07 | -.16 | -.02 | .32 | .42 | .62 | 1.00 | | | | | | | |
| 13. Ready_geometry | .34 | .00 | .18 | .24 | .05 | .09 | -.02 | -.08 | .37 | .48 | .70 | .68 | 1.00 | | | | | | |
| 14. Ready_data | .39 | .09 | .15 | .17 | .10 | .01 | -.18 | -.17 | .26 | .48 | .55 | .65 | .61 | 1.00 | | | | | |
| 15. Professional development | .30 | .30 | .31 | .13 | .29 | .18 | -.16 | -.15 | .41 | .08 | .06 | .38 | .29 | .20 | 1.00 | | | | |
| 16. Class size | .09 | .10 | .05 | .07 | .09 | .16 | .20 | .37 | -.19 | .11 | .16 | .01 | .02 | -.11 | -.07 | 1.00 | | | |
| 17. School resources | -.10 | -.20 | -.02 | -.33 | -.16 | .09 | -.04 | -.16 | -.07 | .15 | -.01 | .06 | .12 | .11 | .19 | -.03 | 1.00 | | |
| 18. Teacher perception_limitation | -.06 | .22 | .00 | -.04 | .24 | .05 | .04 | -.12 | -.19 | .03 | .01 | -.11 | -.08 | -.05 | .02 | .19 | .16 | 1.00 | |
| 19. Math hours per year | .19 | .15 | .06 | .14 | .16 | .02 | -.08 | .11 | .10 | -.14 | -.01 | .09 | -.02 | .16 | .15 | .09 | -.04 | -.14 | 1.00 |

ABOUT THE AUTHOR

Ha Phan received a Bachelor's Degree in English Education from Hanoi University for Teachers of Foreign Languages and a Bachelor's Degree in Finance from National Economics University in Vietnam. At the University of South Florida in the United States, she earned a Master of Education in Instructional Technology. She became interested in measurement and research while pursuing her master's degree and continued with the Ph. D. program in Educational Measurement and Research. During her graduate studies, she worked as a researcher, an instructor, a research consultant, and an editorial assistant for a published journal, the Educational Researcher. Her primary research interest was in the area of assessment and psychometrics. Receiving a fellowship from Harcourt Assessment Inc., in 2007 further increased her passion to work with test scores. Her research has been presented at regional, national, and international conferences. She currently is a research scientist at Pearson Educational Measurement.