Gender, Self-Efficacy, and Mathematics Achievement: An Analysis of Fourth Grade and Eighth Grade TIMSS Data from the United States

Jennifer Anne Evans
Lesley University

Follow this and additional works at: https://digitalcommons.lesley.edu/education_dissertations

Part of the Elementary Education Commons, Gender Equity in Education Commons, and the Science and Mathematics Education Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the Graduate School of Education (GSOE) at DigitalCommons@Lesley. It has been accepted for inclusion in Educational Studies Dissertations by an authorized administrator of DigitalCommons@Lesley. For more information, please contact digitalcommons@lesley.edu, cvrattos@lesley.edu.
Gender, Self-Efficacy, and Mathematics Achievement: An Analysis of Fourth Grade and Eighth Grade TIMSS Data from the United States

A DISSERTATION
Submitted by

Jennifer Anne Evans

In partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Lesley University
School of Education

August 2015
Gender, Self-Efficacy, and Mathematics Achievement: An Analysis of Fourth and Eighth Grade TIMSS Data from the United States

Jennifer Anne Evans

Graduate School of Education
Lesley University

Ph.D. Educational Studies
Interdisciplinary Studies Specialization

Approvals
In the judgment of the following signatories, this Dissertation meets the academic standards that have been established for the Doctor of Philosophy degree.

Dr. Angela Vierling-Claassen
Doctoral Committee Chair

Dr. Brian Becker
Doctoral Committee Member

Dr. Corinna Preuschoff
Doctoral Committee Member

Dr. Caroline Heller
Director, Interdisciplinary Studies Specialization

Dr. Paul A. Naso
Director, Ph.D. Educational Studies

Dr. Jonathon H. Gillette
Dean, Graduate School of Education
Abstract

Gender, Self-Efficacy, and Mathematics Achievement: An Analysis of Fourth and Eighth Grade TIMSS Data from the United States

Dissertation by Jennifer Anne Evans
Advisor: Angela Vierling-Claassen, Ph.D.

It has been argued by some that boys are inherently better in mathematics than girls (Halpern, 2012; Summers, 2005). However, according to international assessments such as the Trends in Mathematics and Science Study’s (TIMSS) and Program for International Student Assessment (PISA), boys do not always outperform girls in mathematics (Mullis, Martin, Foy, & Arora, 2012; OECD, 2014). As such, something other than biology might better explain variations in mathematics performance. One explanation may be self-efficacy, a label used to describe judgments people make about themselves in terms of whether or not they have the capability of doing something (Bandura, 1995; Bandura, 1997). Self-efficacy has been found to have a significant effect on academic achievement (Bandura, 1995; Bandura, 1997; Borman & Overman, 2004; Fast, Lewis, Bryant, Bocian, Cardullo, Rettig, & Hammond, 2010; Pietsch, Walker, & Chapman, 2003).

This dissertation explored the relationship of gender, self-efficacy, and mathematics achievement on the TIMSS assessment as a way to challenge biological arguments that boys are inherently better than girls in mathematics. The country of focus is the United States and the students studied were fourth grade participants who took the 2007 TIMSS test (n = 7,896) and
eighth grade participants who took TIMSS 2011 (n = 10,477). Self-efficacy was examined through responses to selected TIMSS student background questionnaire statements that represented self-efficacy.

Results of this study show that gender on its own is not a significant predictor of mathematics achievement. A positive relationship exists between self-efficacy and mathematics achievement. Further, high self-efficacy is the greatest predictor of mathematics achievement studied in this dissertation. High self-efficacy gave boys a greater advantage in mathematics than girls at both grade levels. This work supports the importance of self-efficacy to mathematics achievement and diminishes the significance of gender to the same end.
To my parents: Philip and Joanne Bargioni
Acknowledgments

The education of children--and adults--“takes a village.” A sincere thank you goes to many people in my life who have helped me come to this point in my education.

First, I would like to acknowledge my committee members for their guidance and support. Dr. Angela Vierling-Claassen gave me critical feedback about my work, cheered me on, and led me to a variety of research articles that helped shaped this dissertation. Dr. Brian Becker made me feel confident that my thinking and work was headed in the right direction and helped me believe that I would graduate when I thought that nothing was further from my reality. Dr. Corinna Preuschoff devoted a great deal of time to the development of this study. She gave me the tools to understand sophisticated quantitative methods and helped to mold my work into a product that I am proud of. I can’t thank her enough for all of her guidance and patience.

Many members of the Lesley University community played an important role in my Ph.D. studies from the start. Dr. Caroline Heller pushed me through the first phase of the program, guided my writing skills and style, and gave me an opportunity to share my ideas about writing research at a conference at UPenn. Fellow graduate students, Mikael Powell, Jibril Solomon, and Claire DiFrancesco coached me through the Ph.D. process and/or edited earlier doctoral works. They inspired me to push on during the most challenging of times.

I am also grateful to Dr. Ina Mullis of Boston College, Executive Director of the TIMSS & PIRLS International Study Center. I will never forget the National Council of Teachers of Mathematics conference in Washington, D.C. where I met and spoke with Dr. Mullis and a few members of the TIMSS team. After this meeting and an email to the TIMSS office in Chestnut Hill, Dr. Mullis agreed to have me work as an intern. I was given the opportunity to help edit
TIMSS reports, input data, and to be a member of a group of people whose efforts resulted in the TIMSS 2011 assessment. I feel blessed to have been a part of such an incredible experience.

I would also like to thank my family and friends who helped me weather the Ph.D. storm. My sister, Christine Carter, sent her unwavering support and care packages from Ohio. My friends, Renee Bernier, Jocelyn Brunner, Liz O’Connell, Megan Fernando, Angela Sheerin, and Erica Willis, made sure that I didn’t forget how to relax and have fun throughout the process. My in-laws, John and Dottie Evans, supported me from the moment they met me. They made me feel like their daughter well before my marriage to their son.

Last, but certainly not least, thank you to two very important people in my life, my husband, Dr. John Warren Evans, III, and our son, John Warren Evans, IV (a.k.a. “Johnny Rockets”). My husband, John, has helped me survive this journey. Toward the end of our programs, you could find us both in the library working on our dissertations in tandem. He offered sound advice, empathy, and reassurance that I would defeat this monster of a task. I’m so glad that we went through this together. And, thank you to “Johnny Rockets.” He was there for the many edits of this dissertation, for my defense, and for the final manuscript. His “love kicks,” which grew stronger over time, often reminded me that there is no greater gift in life than that of a child.

~

This dissertation has been dedicated to my mother and father, Joanne and Philip Bargioni. They’ve always rooted me on, have lifted me up when I’ve needed it, and have instilled in me that I can do anything I set my mind to. The world would be a better place if every child were blessed with parents as exceptional as mine. I’m sure they are not surprised that I’ve earned a Ph.D. They should also not be surprised that it’s because of them.
Table of Contents

CHAPTER 1: INTRODUCTION ......................................................................................................................... 1
  Description of the Problem .......................................................................................................................... 1
    Status of Mathematics Achievement of Boys and Girls on PISA 2012 and TIMSS 2011 ........ 3
  Purpose of the Study ................................................................................................................................. 7
    Research Questions .................................................................................................................................. 8
  Importance of the Study .............................................................................................................................. 8

CHAPTER 2: LITERATURE REVIEW .............................................................................................................. 9
  Gender Construction ................................................................................................................................. 9
  Learning Environments ............................................................................................................................ 12
  Self-Efficacy ............................................................................................................................................. 15
    Factors that Affect Self-Efficacy ............................................................................................................. 20
      Mastery Experience .............................................................................................................................. 20
      Vicarious Experience ............................................................................................................................ 21
    Social Persuasion .................................................................................................................................... 21
    Physiological and Emotional States ....................................................................................................... 22
  Summary ..................................................................................................................................................... 22

CHAPTER 3: METHODOLOGY ..................................................................................................................... 23
  The TIMSS Assessment .............................................................................................................................. 24
    Sampling .................................................................................................................................................. 28
  Mathematics Test Item Development ...................................................................................................... 31
    Mathematics Assessment Framework .................................................................................................... 32
    Item Types ............................................................................................................................................... 34
    Item Blocks ............................................................................................................................................ 35
  Background Questionnaires ..................................................................................................................... 36
    The Reporting and Scaling of Context Questionnaires ........................................................................ 37
  Achievement Means .................................................................................................................................. 38
    Plausible Values ..................................................................................................................................... 38
    Mathematics International Benchmarks ............................................................................................... 39
  This Dissertation’s Methods ..................................................................................................................... 40
    The 2007-2011 Cohort ............................................................................................................................ 40
    Analyses ................................................................................................................................................ 40
List of Tables

Table 3.1. Selected IEA Mathematics Assessments ............................................................... 26

Table 3.2. United States’ 2007 Fourth Grade and 2011 Eighth Grade TIMSS Participants ..... 30

Table 3.3. TIMSS: A Comparison of Mathematics Dimensions and Domains of the Fourth and Eighth Grades ........................................................................................................ 33

Table 3.4. TIMSS 2007: Students Confident in Mathematics Scale (SCM), Grade 4 .......... 42

Table 3.5. TIMSS 2011: Students Confident in Mathematics Scale (SCM), Grade 8 .......... 43

Table 4.1. Factor Analysis of Self-Efficacy Variables of TIMSS 2007, Grade 4................. 47

Table 4.2. Factor Analysis of Self-Efficacy Variables of TIMSS 2011, Grade 8................. 47

Table 4.3. Mathematics Achievement by Self-Efficacy .......................................................... 51

Table 4.4. TIMSS 2007: Fourth Grade Mathematics Achievement by Self-Efficacy and Gender ............................................................................................................................. 54

Table 4.5. TIMSS 2011: Eighth Grade Mathematics Achievement by Self-Efficacy and Gender ............................................................................................................................ 54

Table 4.6. Comparison of Achievement Means by Self-Efficacy and Gender .................... 54

Table 4.7. TIMSS 2007 Grade 4 Regression Models: Relationships between Mathematics Achievement, Gender, and Self-Efficacy ................................................................. 58

Table 4.8. TIMSS 2011 Grade 8 Regression Models: Relationships between Mathematics Achievement, Gender, and Self-Efficacy ................................................................. 58
List of Tables, Continued

Table A.1. TIMSS 2007: List of Participating Countries.................................................................80

Table A.2. TIMSS 2011: List of Participating Countries.................................................................81

Table D.1. TIMSS: 2007 and 2011 Fourth Grade and Eighth Grade Test
Item Block Distribution......................................................................................................................89

Table E.1. Grade 4 and Grade 8 TIMSS Data: Exploratory Analysis of 1st Plausible Values and
Self-Efficacy.........................................................................................................................................91
List of Figures

Figure 1.1. PISA 2012: Mathematics Differences between 15-Year Old Boys and Girls ............ 4

Figure 1.2. TIMSS: Differences in Mathematics Scores between Boys and Girls ................. 6
  Figure 1.2.A. TIMSS 2007 4th Grade Results ................................................................. 6
  Figure 1.2.B. TIMSS 2011 8th Grade Results ................................................................. 6

Figure 4.1. TIMSS 2007: Fourth Grade Self-Efficacy Ratings and Mathematics Means
  By Gender .......................................................................................................................... 55

Figure 4.2. TIMSS 2011: Fourth Grade Self-Efficacy Ratings and Mathematics Means
  By Gender .......................................................................................................................... 55

Figure C.1. TIMSS: Multiple-Choice Item, Fourth Grade ...................................................... 86

Figure C.2. TIMSS: One Point Constructed Response, Eighth Grade ................................. 87

Figure C.3. TIMSS: Two Point Constructed Response, Eighth Grade ................................. 88

Figure E.1. TIMSS 2007: Scatter Plot and Regression Line of 1st Plausible Values and
  Self-Efficacy, Fourth Grade .............................................................................................. 93

Figure E.2. TIMSS 2011: Scatter Plot and Regression Line of 1st Plausible Values and
  Self-Efficacy, Eighth Grade .............................................................................................. 94
Chapter 1: Introduction

Description of the Problem

Gender differences in mathematics performance have been popularized in the United States since the late seventies and early 1980’s (e.g., Benbow, 1980; Halpern, 2012; Tobias, 1976). Magazines such as Newsweek and Time have headlined, “Do Males Have a Math Gene?” (King, 1980) and “The Gender Factor in Math: A New Study Says Males May Be Naturally Abler than Females” (“Behavior,” 1980). In the new millennium, Larry Summers created a press frenzy when he made statements reflecting sentiments that boys were inherently more mathematically inclined than girls. This former Harvard President’s revival of “Boys are better than girls in math” was made, he argued, to explain why women have been underrepresented in mathematically and scientifically based occupations, collectively known as Science, Technology, Engineering, and Mathematics (STEM) (Summers, 2005). He stated that differences in aptitudes, including mathematics, were due to “intrinsic human nature,” and that this intrinsic human nature can be used to explain differences between males and females in areas such as “height, weight, propensity for criminality, overall IQ, mathematical ability, scientific ability” (Summers, 2005). According to Summers, males are genetically more likely to choose a mathematics or scientifically based occupation as a result of “different availability of aptitude at the high end” (2005).

The US Department of Commerce, Executive Office of the President, in cooperation with the Census Bureau, National Center for Education Statistics, and other governmental organizations have investigated why women have not been as likely as men to choose STEM jobs. The US Department of Commerce found that in 2011, women made up nearly half of the
United States work force, but only 26% of total STEM positions (U.S. Department of Commerce, Economics and Statistics Administration, 2013, p.5). Some of the Department’s understandings of why females choose to pursue non-STEM positions include: difficulty balancing childcare and careers, few female role models in STEM, and “…strong gender stereotypes discourage women from pursuing STEM education and STEM jobs” (Beede, Julian, Langdon, McKittrick, Khan, & Doms. 2011, p. 8).

Stereotypes are powerful. Discourse is powerful. Michel Foucault, philosopher, psychologist, and social theorist, has gone so far as to posit that nothing outside of discourse exists (Fendler, 2010; Jardine, 2005). Discourse, including stereotypes, creates conceptual frameworks within and among individuals that may shape identities and establish power structures. These power structures influence the way people interact with each other, verbally or otherwise. For example, a person who views him- or herself in a subordinate position will act as a subordinate in his or her speech or actions. Foucault (1979) refers to this social positioning as a “subject position.” The importance of subject positioning and performance to mathematics teaching and learning is that if students perceive themselves as capable in mathematics, they are more likely to participate during mathematics classes and work hard to do well. The opposite is also true. If students do not think they are good at mathematics, they may not be as engaged in math lessons and may not try as hard to complete math tasks.

Performance and identity, in this case, academic identity, are intrinsically linked (Kurtz-Costes, Rowley, Harris-Britt, & Woods, 2008; Holloway, 2008). Beliefs and perceptions help shape these identities (Davies & Harre, 2008; Holloway, 2008; Kurtz-Costes et al, 2008) and discourse helps mold beliefs. Discourse including stereotypes. Ascribing to certain stereotypes, such as boys are better than girls in mathematics, may transform stereotypical ideas into
individual’s realities. Boys may believe they are better at mathematics than girls and, consequently, perform as such (Davies & Harre, 2008; Foucault, 1979). Of course, while this may be of benefit to some boys, it may also be of detriment to some girls. If girls believe that mathematics is a “boy’s subject” they may socially position themselves, consciously or otherwise, in ways so that they perform, and become, inferior to boys in mathematics (Davies & Harre, 2008; Foucault, 1979).

**Status of mathematics achievement of boys and girls on PISA 2012 and TIMSS 2011.**

According to international assessments, boys’ mathematics achievements are not consistently greater than girls (Mullis, Martin, Foy, & Arora, 2012; OECD, 2014). International data in mathematics show that in some countries: a) boys outperform girls, b) girls outperform boys, or that c) there is no significant difference in mathematics achievement between the two groups (Mullis, Martin, Foy, & Arora, 2012; OECD, 2014).

The Program for International Student Assessment (PISA), an assessment coordinated by the Organization for Economic Cooperation and Development (OECD) and is taken by 15-year old students, present such results (OECD, 2014). In Figure 1.1, PISA countries are rank-ordered so that the countries where girls outperformed boys by the greatest margins are listed at the top and countries where boys outperformed girls by the greatest margins are listed at the bottom.
Figure 1.1. PISA 2012: Mathematics Differences between 15-Year-Old Boys and Girls

Notes.
1. Statistically significant differences are marked in the darker tone.
2. Figure adapted from *PISA 2012 results: What students know and can do* (OECD, 2014, p. 73).
Trends in Mathematics and Science Study (TIMSS) 2011 data deliver similar results in which there are countries where boys outperform girls, girls outperform boys, or no differences exist between the sexes.

Figures 1.2.A. and 1.2.B. display international mathematics results from TIMSS 2007 and 2011 of a cohort of students\(^1\). Countries listed within these figures have been ordered by differences in mathematics achievements of girls and the boys. Countries with the least differences have been placed at the top of the figures and continue downward toward countries with the greatest differences. Each figure is split in half. Bars that extend to the left of the division show that females scored higher than boys. Bars that extend to the right of the division, show that males had higher means than girls. In looking at these two figures, it can be seen that in the fourth and eighth grades, boys do not routinely outperform girls. Additionally, there are a number of countries in which females perform better than boys. This is particularly true in 2011 for the eighth grade TIMSS test.

\(^1\) Students within this cohort either took the fourth grade TIMSS in 2007 or the eighth grade TIMSS in 2011. TIMSS is not a longitudinal study.
Figure 1.2: TIMSS: Differences in Mathematics Scores between Boys and Girls

Figure 1.2.A: TIMSS 2007 4th Grade Results

<table>
<thead>
<tr>
<th>Country</th>
<th>Gender Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Girls Scored Higher</td>
</tr>
<tr>
<td>Lithuania</td>
<td></td>
</tr>
<tr>
<td>England</td>
<td></td>
</tr>
<tr>
<td>Ukraine</td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td></td>
</tr>
<tr>
<td>New Zealand</td>
<td></td>
</tr>
<tr>
<td>Chinese Taipei</td>
<td></td>
</tr>
<tr>
<td>Latvia</td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td></td>
</tr>
<tr>
<td>Morocco</td>
<td></td>
</tr>
<tr>
<td>Hungary</td>
<td></td>
</tr>
<tr>
<td>Hong Kong SAR</td>
<td></td>
</tr>
<tr>
<td>Algeria</td>
<td></td>
</tr>
<tr>
<td>Slovenia</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td></td>
</tr>
<tr>
<td>Czech Republic</td>
<td></td>
</tr>
<tr>
<td>Singapore</td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td></td>
</tr>
<tr>
<td>Slovak Republic</td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td></td>
</tr>
<tr>
<td>Norway</td>
<td></td>
</tr>
<tr>
<td>Russian Federation</td>
<td></td>
</tr>
<tr>
<td>Kazakhstan</td>
<td></td>
</tr>
<tr>
<td>Armenia</td>
<td></td>
</tr>
<tr>
<td>Scotland</td>
<td></td>
</tr>
<tr>
<td>El Salvador</td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td></td>
</tr>
<tr>
<td>Iran, Islamic Rep. of</td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td></td>
</tr>
<tr>
<td>Colombia</td>
<td></td>
</tr>
<tr>
<td>Tunisia</td>
<td></td>
</tr>
<tr>
<td>Yemen</td>
<td></td>
</tr>
<tr>
<td>Qatar</td>
<td></td>
</tr>
<tr>
<td>Kuwait</td>
<td></td>
</tr>
</tbody>
</table>

Notes.

1. Adapted from TIMSS 2007 international mathematics report: Findings from IEA’s trends in international mathematics and science study at the fourth and eighth grades (Mullis, Martin, & Foy [with Olson, Preuschoff, Erberber, Arora, & Galia], 2008, Exhibit 1.5, p. 58).
The PISA and TIMSS results discussed are not new. Though the information in Figures 1.2.A. and 1.2.B. were published after the 2001-2006 Summers reign, TIMSS data in 2003 (Mullis, Martin, Gonzalez, & Chrostowski, 2004, p. 49-50) and PISA data in 2003, 2006, and 2009 (Baldi, Jin, Green, & Herget, 2007, p. 14; Lemke, Sen, Pahlke, Partelow, Miller, Williams, Kastberg, & Jocelyn, 2004, p. 36) have also provided evidence that boys don’t always outperform girls in mathematics. In fact, the figures listed under 1.2 show that a higher percentage of girls outperformed boys on the eighth grade TIMSS 2011 than on the fourth grade TIMSS 2007. Not only does this suggest a shift in who has a mathematics advantage, boys or girls, it also suggests that something other than biology might better explain the differences in mathematics performances, or lack of, between the two groups.

**Purpose of the Study**

As indicated by PISA and TIMSS, gender advantages in mathematics performance aren’t consistently achieved by males. International data shows that females outperform males, in other cases, that no statistical differences appear between the two groups. There is also evidence that there is an increase in females outperforming males on the eighth grade TIMSS of 2011 than on the fourth grade TIMSS of 2007. In the United States, although boys outperformed girls in mathematics on both of these assessments, the achievement gap between the boys and girls of this cohort decreased from 2007 to 2011 (Mullis, Martin, Foy, & Arora, 2012; Mullis, Martin, & Foy [with Olson, Preuschoff, Erberber, Arora, & Galia], 2008). Male biological advantages in mathematics seem to be challenged by this data. So, what factors may be associated with this shift in mathematics achievement between boys and girls?
One possible explanation for this variation in mathematics performance is the belief students have in their mathematics abilities, self-efficacy. According to the father of self-efficacy, Albert Bandura, self-efficacy has a significant effect on academic achievement (Bandura, 1995; Bandura, 1997). Given boys and girls may outperform each other on international assessments, there seems to be something else other than biology that may better explain students’ mathematics achievements.

It is with this framework in mind that I examined the mathematics achievements of boys and girls in the 2007-2011 United States cohort.

Research questions.

Research questions steering the efforts of this dissertation are:

1. Is there a relationship between gender, self-efficacy, and mathematics achievement?

2. How well do gender and self-efficacy predict mathematics achievement?

Importance of the Study

The heart of this dissertation is equity. Statements of male superiority in math used to explain a lack of females in the STEM field reflects beliefs of inherent inequity; something that can’t be avoided and may yield greater achievement gaps between males and females.

Investigating the self-efficacy of students, gender, and mathematics performance on the TIMSS test may shed light on how students’ self-efficacy may or may not influence mathematics achievement. It offers a different explanation as to why there may or may not be mathematics differences between males and females. A study like this is important in terms of contributing information to the field of mathematics education, particularly as it relates to pedagogy and equitable access to academic success and pursuits.
Chapter 2: Literature Review

Decades of research have been devoted to why sex differences in mathematics exist (Fennema & Sherman, 1978; Keller & Menon, 2009; Stoet & Geary, 2013). In this chapter, I review areas of research related to constructions of identity and social influences as a way to explain sex differences in mathematics. First, I explain the construction of gender. Then, I discuss social influences within learning environments. Finally, I write about Bandura’s Theory of Self-Efficacy and its relationship to mathematics performance. These theories and points of discussion have been included in chapter 2 to lay a bedrock for methods decisions described in Chapter 3.

Gender Construction

“‘No, no no, I am a boy, a boy, a boy’ he said…At this point Michael pulled down his pants, pointed to his genitals and exploded with obvious impatience ‘Here, look, I am a Boy!’” (Davies, 1989, p. 237). This quote is a reaction given by a four-year-old boy returning to his preschool with a note from his father saying he should no longer put nail polish on at the school. The day before, Michael applied the nail polish to his nails with enjoyment. The next day, however, he returned to school telling his teachers he was a “good boy” and “boys don’t wear nail polish” (p. 237). When his teachers spoke to him about the event and agreed that boys don’t wear nail polish but “some of the boys here like to try things when they play that are different, just for fun” (p.237). The boy exploded with the above quote to prove his gender to his instructors.

This incident illuminates how gender is not biologically defined; it is socially constructed. If gender was defined by anatomy alone, there would be no need for Michael’s father to write a note. Michael’s gender would already be defined. However, masculinity and femininity are not static. They are constructed and reconstructed through a variety of social interactions. As stated in Davies’ (1989) article, “the individual is not so much the product of some social process of social
construction that results in some relatively fixed end-product but is constituted and reconstituted through the various discursive practices in which they participate” (p. 229). As people witness, discuss, and live these experiences, they assimilate the information within their own belief systems and then act (i.e., speak, dress, or behave) accordingly.

Gender construction begins early on in life. Family and friends begin the process through gifts (such as a truck for a boy, a doll for a girl for a child’s birthday, or clothes in pink or blue) (Halpern, 2012, p. 292-293). The media also plays a role in defining what is feminine and what is masculine through music, clothes worn by actors and actresses, script development, etc. (p. 293). Gender construction exists everywhere in a society and is often unknowingly transmitted to and through people. This unconscious gender practice and maintenance is coined by some researchers as “nonconscious ideology” to label a “tendency to underestimate [tacit] sex-differentiated experiences and messages, and expectations.” People socialize and are socialized as “fish that are unaware that the water is wet” (p. 256). The fish are limited by their experience. The same is true for human beings. It’s difficult for people to recognize something like gender construction and maintenance because they are completely immersed in it.

In one preschool study (Anggard, 2005), a small group of children was asked to create books. They had the liberty to choose any subject. When the teacher went from student to student to find out what they intended on making, the boys stated they wanted to create stories about dinosaurs, dragons, or knights. And, when the teacher came to Jenny, the only girl present that day in class, to ask her what her story would be about, the boys in her class emphatically advised that she write about Barbie, Barbie princesses, or a king with Barbie (p. 543). In the end, she conceded to the boys’ suggestions and ended up creating a story about Barbie. And, when the
girls who were absent from preschool returned the next day, they looked at Jenny’s work and decided to write about Barbie, too.

The children of this class all decided to create gender-flagged stories. The boys developed stories about heroes and adventures, and the girls made stories about Barbie. This highlights the idea that children may socially influence each other at a very young age and, as is captured from Davies and Harris’s piece in Anggard, that “every narrative about ourselves must be understood as negotiated collaboratively in joint action” (Anggard, 2005, p. 543).

Gender pressure certainly extends beyond that of preschool walls and coercion by people within immediate social circles. Pop culture and media are part of this socialization process, too. Gebauer and Wulf (1995) describe a social theory, mimesis, in which people mimic the social world (p.10). This mimicry is not passive nor is it simplistic (p. 11). Mimetic approaches allow people, especially those working on the formation of identity, such as children and adolescences, to navigate social expectations through their experiments with mimicked speech, interests, behaviors, and dress (Fritzsche, 2004, p.157)

An example of this can be found through a case study of a girl, Julia, a self-proclaimed fan of a once popular all-girl band called, The Spice Girls (Fritzsche, 2004). At the age of 13, she found comfort in being a part of the Spice Girls’ fan culture as she was “shy and incapable of fighting for her own interests” (p.157). She stated, “[G]radually, one grew into the role. We became more and more similar to them” (2004, p. 158). She and her friends created dances, wore outfits in the Spice Girls’ style, and adopted the attitudes and expressions of the music group. Julia mimicked the speech and behaviors of the most outgoing member, Crazy Spice. When she was Crazy Spice, Julia would act confidently—though ordinarily she was not. Mimicry of the
behaviors and Julia’s perceived attitudes of the Spice Girl, Crazy Spice, enabled her to experiment with a self-confidence she didn’t possess at the time herself.

This mimicry also shaped how she negotiated her relationships with boys. “We need boys as lovers, but otherwise forget about them, don’t run after them, find your own way instead” (Fritzsche, 2004, p. 158). Using the Spice Girls’ attitude and lyrics to pave her way, she started objectifying boys. She believed a boy’s existence to be near useless and only meant for primitive ends. However, as time passed, she distanced herself from her Crazy Spice identity and lifestyle as her beliefs started to change. As she interacted more with boys, she started to recognize boys as human not just things to be used. Her Spice Girls and real world beliefs started to clash more and more until eventually she completely abandoned her Spice Girls persona for one that resonated better with her new found beliefs.

The use of “transmedia intertextuality,” a phrase coined by Marsha Kinder (1991), in which t-shirts, games and toys replicate popular media, such as Spice Girls products, help to create a strong force in persuading children and adults to assimilate sensationalized ideas or norms. Young people are susceptible to the adoption of beliefs, behaviors, and ideas presented to them via the media or through their conversations with others. Through these means, they, like Julia, may be coerced into promoting and displaying what it means to be feminine (and/or masculine) even if it’s in spite of their interests or natural tendencies.

**Learning Environments**

Beliefs, including those involving equity for all people, begin as thought and are part of social processes (Vygotsky, 1986). For children, many hours of their day are spent interacting with teachers and peers. Educational organizations like the National Council of Teachers of
Mathematics (NCTM) and Achieve, the group responsible for the development of the Common Core Standards, have made sure to promote philosophies of equity to promote equal access and opportunities for all children of the United States of America. One of the NCTM’s founding principles states that all teaching and learning environments should foster “a culture of equity where everyone is empowered by the opportunities mathematics affords” in the United States (2000, para. 6). And, Achieve’s curriculum standards, adopted by 43 states, the District of Columbia, Puerto Rico, Guam, American Samoan Islands, US Virgin Islands, Northern Marina Islands, and the Department of Defense Education Activity (Closing the expectations gap, 2013), have been “[i]nformed by other top-performing countries to prepare all students for success in our global economy and society” (Read the Standards, 2015, para. 2).

Nevertheless, interactions between students and teachers are laden with language and actions saturated by biases and stereotypes (Helwig, Anderson, & Tindal, 2001) that may create social inequities in classrooms. For all speech begins as thought (Vygotsky, 1986), and as such, social interactions can never be devoid of belief systems consciously or subconsciously held.

Teachers, of course, use language to instruct, question, check student understanding, establish relationships with students, and, as neutral as they may try to be, they cannot completely guard students from the their biases (Helwig, Anderson, & Tindal, 2001). If teachers believe that boys are generally better at mathematics than girls (or the inverse), it may be picked up by their students by the difficulty of the questions teachers pose or the types of activities that they choose for some students instead of others.

One research pair, Robert Rosenthal and Lenore Jacobson (1968), developed a study on teacher biases. In their study, they told teachers in a public elementary school certain children were likely to have enhanced academic performances, which they named “growth spurt,” based
on Harvard Test of Inflected Acquisition. In truth, the Harvard test was fictitious and the children were randomly selected for “pretend” growth spurts. The results of the study showed that the children who were identified as growth spurt candidates had improved academic performance. Rosenthal and Jacobson explained that this may be because teachers unconsciously behaved in ways that may have supported (or not) the students’ mathematics learning.

Fast-forwarding, Joachim Tiedemann (2000) conducted research in fifty elementary schools. In her study, mathematics teachers were asked questions about the achievements of boys and girls. The results of the study showed that teachers believed their average performing girls were less skilled in mathematics than average performing boys, that girls would not benefit from exerting additional effort to the extent boys would for teachers, and that mathematics was rated by teachers as being more difficult for average achieving girls than boys equally achieving (Tiedemann, 2000).

In a study by the University of Chicago, an attitudinal survey was given to seventeen first- and second-grade female elementary school teachers at the beginning of the school year (Beilock, Gunderson, Ramirez, & Levine, 2010). At the beginning of the school year, there was “no relation” between the teachers’ anxieties and mathematics performances of the students (52 boys and 65 girls). However, by the end of the school year, the more anxious teachers were about mathematics themselves, the more likely their female students would endorse stereotypes such as “boys are good at math and girls are good at reading.” This study also found that females who believed in that stereotype performed significantly lower on a variety of tests than females who didn’t.

Schools harness a power to reproduce or shape norms (Foucault, 1980). The activities that take place within schools have the ability to change discourse and thereby change students’
beliefs about power and identity. The primary vehicle for this is the teacher. Teachers have a
direct ability to change discourse and change stereotypes with regard to much, including who’s
good at mathematics and who is not. They have the ability to improve how children feel about
themselves, and consequently improve student achievement (Bandura, 1995). Status quos can be
changed if a change in discourse takes place, for a shift in perception requires a shift in speech
(Edley, 2001).

Even as unethical as Rosenthal & Jacobson’s methods were, the Rosenthal & Jacobson study showed that teachers profoundly affect how students view themselves as learners and, consequently, their academic performance.

And, again, in the Beilock, Gunderson, Ramirez, & Levine (2010) study, teachers’ discourse influenced behaviors. In this case, when the classroom discourse supported stereotypes that “boys are better in mathematics” and “girls are better at reading,” what was prophesized came true.

**Self-Efficacy**

When a social environment permeates the psyche of students, students’ identities may be influenced and their behaviors or attitudes altered. One psychologist who has spent his professional life describing identities and emotions as they relate to action and motivation is Albert Bandura. He has devoted his life’s work to his theory of self-efficacy. Self-efficacy describes the judgments people make about themselves, in terms of whether or not they have the capability of doing something, and how likely they feel that such an outcome will happen (Bandura, 1995; Bandura, 1997).
What’s most interesting about his social theory of self-efficacy is the idea that self-beliefs can empower people to overcome obstacles, social or otherwise, and result in academic achievement (Bandura, 1993, p. 118; Pajares & Urdan, 2006). People with high self-efficacy are more likely to work harder, reflect on their progress more frequently, and are more likely to keep working until they find strategies that will help them to be successful (Schunk & Pajares, 2005).

Self-efficacy has been looked at as a predictor of academic achievement for years. Evidence has shown that individuals who have higher self-efficacy in a subject demonstrate higher academic achievement in that subject area (Borman & Overman, 2004; Fast, Lewis, Bryant, Bocian, Cardullo, Rettig, & Hammond, 2010; Pietsch, Walker, & Chapman, 2003). Pajares and Miller (1994) have provided evidence that mathematics self-efficacy is a greater indicator of mathematics performance than previous mathematics experiences.

Fast et al. (2010), in a study of 1,163 elementary school children from low to middle incomes from southern California, examined how children felt about their classroom environments (i.e., how caring or challenging it felt) and compared these feelings to self-efficacy and mathematics performance (p. 729). Self-efficacy was measured by the Student Motivation Questionnaire (SMQ), a questionnaire developed by researchers on the National Science Foundation-funded Math-Science Partnership—Motivation Assessment Project (p. 731). The SMQ also measured students’ perceptions of how much their teachers emphasized hard work and an intrinsic value of mathematics (e.g., survey item: “My teacher thinks it’s important to understand our math work, not just memorize it”), perceptions of academic rigor (e.g., survey item: “Our math teacher doesn’t let us get away with doing easy work but really makes us think”), and perceptions of teacher’s caring (e.g., survey item: “Our math teacher takes a personal interest in students”). Students rated each item using a 5-point scale. Mathematics performance was
measured by the California Standardized Test (CST) results of the students (Fast et al., 2010, p.732). The results of the study showed that self-efficacy was positively related to students’ perceptions of teachers and mathematics performance (p. 734).

People may find sources of agency from positive self-beliefs. Bandura argues, “Among the mechanisms of agency, none is more central than people’s beliefs about their capabilities to exercise control over their own level of functioning and over events that affect their lives” (Bandura, 1993, p.118).

Bandura (1997) is careful to make a distinction between self-efficacy and confidence. He argues that they are not the same.

It should be noted that the construct of self-efficacy differs from the colloquial term "confidence." Confidence is a nondescript term that refers to strength of belief but does not necessarily specify what the certainty is about. I can be supremely confident that I will fail at an endeavor. Perceived self-efficacy refers to belief in one's agentive capabilities, that one can produce given levels of attainment. A self-efficacy assessment, therefore, includes both an affirmation of a capability level and the strength of that belief.

Confidence is a catchword rather than a construct embedded in a theoretical system. (Bandura, 1997, p. 387)

Confidence is not domain specific. People may be confident that they are good students even though they may also identify that mathematics is difficult for them. Confidence is an overall feeling and self-efficacy is a form of feeling compartmentalization. That is, people may view themselves as good students overall (confidence) even if they perceive that they struggle with one subject area, e.g., mathematics (self-efficacy). Connectedly, those people may have a low self-efficacy in mathematics, and higher self-efficacies in the other subjects.
Additionally, self-efficacy does not just describe the act of compartmentalizing academic subjects. It also describes an outcome of that belief—specifically, the belief in the ability to complete a goal or task (Bandura, 1997).

With the framework that thought, including self-beliefs, is a precursor to action, Bandura explains that beliefs steer the course in which action takes place (Bandura, 1997, p. 118). People first establish belief systems about themselves and act in accordance. For those who believe they are failures in a particular area, thus they will perform. The opposite is also true. Additionally, the stronger the self-efficacy, the harder people will work to achieve those goals (p. 118). Personal goal setting is born within a positive self-efficacious mind. Bandura continues with “[t]he major function of thought is to enable people to predict events and to develop ways to control those events that affect their lives. Such skills require effective cognitive processing of information that contains many ambiguities and uncertainties” (Bandura, 1997, p. 120). People who are riddled with self-doubt produce fewer changes to improve their situation in relation to those with higher self-efficacy whose firm belief systems provide greater ingenuity and resilience to overcome obstacles (p. 125). Bandura found that,

Initially, people relied heavily on their past performance in judging their efficacy and setting their aspirations. But as they began to form a self-schema concerning their efficacy through further experience, their performance attainments were powered more strongly and intricately by their belief in their personal efficacy.

(p. 128)

Perceived control is at the helm of success and failure.

Additionally, self-efficacy can also influence a person’s affect. If confronted with difficult situations, the capability of overcoming obstacles can be impeded or bolstered by a low or high
self-efficacy, respectively. People who have higher self-efficacies are less likely to experience stress or depression about a situation because they feel that they can manage the problem. Self-efficacy has the ability to influence stress and anxiety (Bandura, 1997, p. 132). It can act as an emotional control, it may also entice people to make bolder choices and try more aggressive solutions to their problems (Bandura, 1997, p. 135). With the belief that impediments can be overcome, a greater variety of activities and actions may be taken or explored and as a result, more interests may be cultivated. For individuals with lower self-efficacy, the opposite may be true (p.135). Low self-efficacy may promote an avoidance of behaviors or activities that are is viewed as too difficult to overcome.

Hackett and Betz (1989) studied the relationships between mathematics self-efficacy, mathematical performance on the America College Test (ACT), and college major choice of 153 female and 109 male undergraduates. Using the Fennema-Sherman Mathematics Attitudes Scales (MAS) to measure mathematics self-efficacy, Hackett and Betz confirmed a correlation between mathematics performance on the ACT and mathematics self-efficacy (p. 265). They found that mathematics self-efficacy was also a strong predictor of a college major choice related to mathematics (p. 269).
Factors that affect self-efficacy.

There are four factors that Bandura has identified to affect the self-efficacy of an individual. They are:

1. mastery experience, the past performance of an individual
2. through vicarious experiences, observing others
3. social persuasion that one “can do it”
4. physiological and emotional states

(Bandura, 1986).

Mastery experience.

Of the four factors, mastery experience, or past experience, is the greatest contributor to self-efficacy (Bandura, p.3, 1995; Usher & Pajares, 2006, p. 7). Once a task has been completed, people evaluate their experience and think about what went well or techniques that might help future tasks be completed with greater ease. This information helps inform future successes and may increase self-efficacy. When people have previously been successful on similar or identical tasks, they are more likely to believe they will succeed on comparable tasks (Bandura, 1993; Bandura, 1995). Alternatively, if people have failed tasks, their self-efficacy may be diminished when attempting similar undertakings (Bandura, 1995; Usher & Pajares, 2006). Success and failure of a task is determined by an individual’s interpretation of the event. Interpretations may be shaped by perceived difficulty of the material, amount of assistance received, amount of effort required of the individual, and, in terms of academic results, perceived level of subjectivity of the grade or score (Bandura, 1997).
**Vicarious experiences.**

The second listed influence of self-efficacy, vicarious experience, describes what an individual internalizes as a result of observing others’ experiences. By observing the successes or failures of other people trying to complete a task or assignment, an individual might calibrate the likelihood of his or her own potential capabilities (Bandura, 1997; Bandura, 1986). It may also prove to someone with low self-efficacy, that he or she has the ability to be successful at the task, too (Bandura, 1997).

The impact of the vicarious experience to the individual absorbing the information is also dependent on perceived similarity. The more similar a person is to the observer, in term of the perceived similarities in skill sets required the complete the task, the more likely a vicarious experience will effect an individual (Bandura, 1997, p. 3). Moreover, modeled influences provide more than just a point of calibration for the observer. They may also transmit knowledge, strategies, and skills that could be used to accomplish a task and increase the observer’s self-belief that he or she can also complete a task with success.

**Social persuasion.**

Another source of self-efficacy is social persuasion. Social persuasion describes when others influence a person’s belief that he or she can achieve something because he or she possesses the abilities needed to complete a job (Bandura, 1997, p.4). The more influential or significant a person is to the individual, the greater the self-efficacy gain (Usher & Pajares, 2006). Additionally, if these people have proved that they were right in the past in terms of an individual’s capabilities, the greater the persuasive force they will have (Bandura, 1997). For students, these influential people may be teachers, parents, and peers.
Physiological and emotional states.

Lastly, Bandura proposes that people’s self-efficacy can be impacted by their emotional states (i.e., anxious, excited, or calm) and physiological states (i.e., fatigue or sweaty palms and rapid heartbeat due to stress or nervousness) (Bandura, 1986). These signs may undermine the self-efficacy beliefs of an individual, or in the case of more positive feelings like excitement, bolster it. People gauge their capabilities by interpreting how these feelings indicate their potential degree of success or failure (Bandura, 1997, p. 7). Feelings of stress or anxiety about mathematics, for example, could undermine their mathematics self-efficacy. Emotional or physical responses of people may be interpreted as a lack of ability to succeed (p. 7).

Summary

This chapter discussed gender construction, learning environments, and Bandura’s theory of self-efficacy to explain sex differences in mathematics and factors that affect student performance. The social and emotional environments students learn in influence teaching and learning. All people live and socialize through a viscous social media of discourse, commercialism, and social punishments and rewards. As neutral as people may try to be or think they are, their identity and belief systems are built on the shoulders of tacit norms shaped by social feedback; feedback that has been shown in this chapter to have a relationship with mathematics performance and self-efficacy.

In the third chapter, I discuss methods used to further study the role of gender and self-efficacy and to mathematics achievement.
Chapter 3: Methodology

The goal of this dissertation is to examine the relative difference in mathematics achievement between boys and girls in the United States and how this difference changes as students move up in grade levels. To answer this question, this dissertation examines the average mathematics scores of fourth grade students who took the TIMSS test in 2007 and eighth grade students who took TIMSS test in 2011, both being part of the same cohort. Since students in this 2007-2011 cohort are representative samples of the same population, the two populations are compared. Furthermore, this analysis may provide information about factors that may have contributed to gender differences in this population of students, specifically self-efficacy, as it is measured by TIMSS student background questionnaire data. To do this, first, I examined gender differences within the same testing year and grade level. Then, I analyzed the factors, which may explain these differences by looking at the relationships between gender, self-efficacy, and mathematics achievement at both grade levels relative to each other.

This chapter is divided into two main sections. In the first section, I briefly discuss what TIMSS is and its development from its beginnings to 2011. Next, I describe TIMSS sampling procedures and how these procedures produce samples representative of the population in participating countries. Later, I discuss how mathematics achievement and students’ self-efficacy in mathematics are measured in TIMSS. In the second section, I discuss my analysis. That is, how I validated the self-efficacy scale and how I investigated gender differences as well as the relationships between gender, self-efficacy, and mathematics achievement.
The TIMSS Assessment

The Trends in International Mathematics and Science Study (TIMSS) has measured trends in student progress in mathematics and science in countries from around the world since 1995 (Martin, 1996). Policy makers, educational researchers, and others have used TIMSS data to garner information about students’ science and mathematics achievement. Every four years, the assessment is given and may be used to examine student growth, achievement, and achievement differences between population subgroups. Such information has the potential to lend itself to a better understanding of how children learn best and the sculpting of educational practice and policy.

Though currently managed by the TIMSS & PIRLS International Study Center at Boston College (Mullis, Drucker, Preuschoff, Arora, & Stanco, 2012, p.7), TIMSS was founded by the International Association for the Evaluation of Educational Achievement (IEA), an international assessment group with offices in Germany and Amsterdam. The seeds of TIMSS were planted by an interest to obtain data “from across a wide range of systems, the variability would be sufficient to reveal important relationships, which would otherwise escape detection within a single education system.” IEA believed this kind of work, “strongly reject[ed] data-free assertions about the relative merits of various education systems, and [may help to] identify factors that would have meaningful and consistent influences on educational outcomes” (A brief history of IEA: 50 years of educational research, n.d., para. 2).

En route to TIMSS was the Pilot Twelve-Country Study designed to determine whether or not a large international study would be possible in 1960 (para. 3). This study focused on mathematics, reading comprehension, geography, science, and a “non-verbal” ability of 13-years-olds (para. 4). It yielded useful education data, but more significantly, the actual design and
implementation of a multi-country assessment proved feasible and led to more involved international comparisons. In mathematics, this included the First International Mathematics Study (FIMS), Second International Mathematics Study (SIMS), and the Third International Mathematics and Science Study (TIMSS 1995), now known as the Trends in International Mathematics and Science Study (TIMSS). With each successive assessment, improvements were made in terms of sampling procedures, number of participating countries, grade levels being assessed, and eventually included background questionnaires (e.g., questions about parent education, teacher education, time spent on homework, interest in mathematics and science, etc.) used to understand the educational context of students and to examine correlations between environmental factors and student mean scores (*A brief history of IEA: 50 years of educational research*, n.d.). Table 3.1 shows the evolution and development of IEA studies in mathematics from 1960 to 2011.
Table 3.1. Selected IEA Mathematics Assessments

<table>
<thead>
<tr>
<th>Year</th>
<th>Name of Selected IEA Assessment</th>
<th>Number of Countries Assessed</th>
<th>Students Tested</th>
<th>Areas Assessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>Pilot Twelve Country Study</td>
<td>12</td>
<td>13-year-olds</td>
<td>mathematics, reading comprehension, geography, science, and non-verbal ability</td>
</tr>
<tr>
<td>1964</td>
<td>First International Mathematics Study (FIMS)</td>
<td>12</td>
<td>13-year-old and final-year secondary students</td>
<td>mathematics</td>
</tr>
<tr>
<td>1980–1982</td>
<td>Second International Mathematics Study (SIMS)</td>
<td>20</td>
<td>13-year-old students and students in the final year of secondary school</td>
<td>mathematics</td>
</tr>
<tr>
<td>1995</td>
<td>Third International Study of Science and Mathematics (TIMSS 1995)</td>
<td>46</td>
<td>9-years-olds, 13-year-olds (tested grades with highest percentage of the above age groups), and students in their final year of secondary school</td>
<td>mathematics &amp; science</td>
</tr>
<tr>
<td>1999</td>
<td>Third International Mathematics and Science Study (TIMSS 1999; also known as TIMSS-Repeat)</td>
<td>38</td>
<td>eighth grade</td>
<td>mathematics &amp; science</td>
</tr>
<tr>
<td>2003</td>
<td>Trends in Mathematics and Science Study (TIMSS 2003)</td>
<td>51</td>
<td>fourth grade &amp; eighth grade</td>
<td>mathematics &amp; science</td>
</tr>
<tr>
<td>2007</td>
<td>Trends in Mathematics and Science Study (TIMSS 2007)</td>
<td>62</td>
<td>fourth grade &amp; eighth grade</td>
<td>mathematics &amp; science</td>
</tr>
<tr>
<td>2011</td>
<td>Trends in Mathematics and Science Study (TIMSS 2011)</td>
<td>64</td>
<td>fourth grade &amp; eighth grade</td>
<td>mathematics &amp; science</td>
</tr>
</tbody>
</table>

Notes. 1. Benchmarking participants are not counted separately from their country under “Number of Countries Assessed.” For example, in 2011, nine states in the United States of America were counted as 1 under “Number of Countries Assessed” since those states are within one country.
2. Adapted from *A brief history of IEA: 50 years of educational research.* (n.d.).
The IEA organization went from assessing 12 countries in 1960 to over 60 countries in 2011 (Mullis, I.V.S. & Martin, 2011a). Additionally, regions within countries, such as Massachusetts in the United States of America, and Alberta in Canada, were also able to participate in the assessment. They are known as benchmarking participants.\footnote{See Appendix Table A.1 and Table A.2 for more information about benchmarking participants.}

The advantage of taking the TIMSS test for any country or benchmarking participant is that it provides a means to compare educational systems (i.e., educational structure, curricula, and pedagogy) with that of other countries and may arm a country with information that may be used to change education policy and improve student learning (Mullis & Martin, 2011a, para. 2).
Fourth and eighth grade students have been assessed during the same testing year since TIMSS 2003\(^3\) (Martin & Mullis, 2004; Martin & Mullis, 2008; Martin & Mullis, 2012; Mullis, Martin, Ruddock, O’Sullivan, and Preuschoff, 2009). The reason behind this decision was that the fourth grade concludes 4 years of primary school and the eighth grade is the end of lower-secondary education (Mullis et al., 2009, p. 13). Collecting data from the fourth and eighth grades helps assess what is happening at the end of two important educational phases of a child’s formal schooling leading up to twelfth grade. Differences may then be examined between the fourth and eighth grade of the same year or differences across time.

**Sampling.**

School and student selection for participation in TIMSS requires several steps. The objective of the sampling process is to select a population of students that is representative of the demographics of the population in a given country.

In the fourth grade, all students must have completed 4 years of formal schooling and have at least a mean age of 9.5 years old at the time of testing to participate; target ages are based on United Nations Educational, Scientific and Cultural Organization’s (UNESCO)\(^4\) International

\(^3\) Only fourth and eighth grade students have been assessed through TIMSS every four years since 2003. However, in 1995, students were tested in the third, fourth, seventh, eighth grade, and final year of secondary (Mullis, Martin, Gonzalez, Gregory, Garden, O’Connor, Chrostowski, & Smith, 2000, p. 313). In 1999, only eighth grade students were tested to measure trends in achievement from 1995 (p. 313).

\(^4\) An agency of the United Nations
Standard Classification of Education (ISCED) (Joncas, 2008a, p. 78, chapter 5). In the eighth grade, all students must have finished eight years of formal schooling and their mean age should be 13.5 years old (Joncas & Foy, 2012, p. 4). Specific information about 2007 fourth grade and 2011 eighth grade students from the United States are listed in Table 3.2.
### Table 3.2. United States’ 2007 Fourth Grade and 2011 Eighth Grade TIMSS Participants

<table>
<thead>
<tr>
<th>TIMSS Year</th>
<th>Grade</th>
<th>Mean Age</th>
<th>School Sample Size</th>
<th>Classes Sampled Per School</th>
<th>Student Sample Size</th>
<th>Stratifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>4</td>
<td>10.3 years old</td>
<td>257</td>
<td>2</td>
<td>7,896</td>
<td>school type (public, private), geographic location (northeast, southeast, mid-west, west), and minority status</td>
</tr>
<tr>
<td>2011</td>
<td>8</td>
<td>14.2 years old</td>
<td>501</td>
<td>1</td>
<td>10,477</td>
<td>high/low poverty index, school type (public, private), and geographic location (northeast, southeast, mid-west, west), degree of urbanization (city, rural, suburban, or town), and ethnic status</td>
</tr>
</tbody>
</table>

**Notes.**


2. Classes sampled in the eighth grade are just one per school because of a national study taking place in the United States. One classroom in a selected school took the TIMSS and the other classroom took the national assessment (TIMSS 2011 characteristics of national samples, 2012, p. 106).
Mathematics test item development.

In mathematics and science, TIMSS participants are tested on questions based on assessment frameworks developed by a committee from participating countries (Martin & Mullis, 2012, p. 1-4). All proposed test items go through a rigorous evaluation by the TIMSS Science and Mathematics Item Review Committee (SMIRC) (Martin & Mullis, 2012, p. 2).

The teams meet 4 times within the 4-year cycle:

1. To review the content frameworks that scaffold the test items’ design,
2. Write and review field test items before and after implementation,
3. Write and review the scoring of the questions and the scaling of scores, and
4. To review and describe the reporting scales\(^5\).

(Martin & Mullis, 2012, p. 2)

The recommendations of SMIRC then go to the TIMSS National Research Coordinators (NRC), who represent all participating countries, to evaluate and revise the frameworks and to test questions and reporting scales.

\(^5\) Scales, in this case, refer to constructs being tested. For example, the students’ confidence in mathematics (SCM) scale in the fourth grade was tested through context questionnaire items on TIMSS 2011 such as “I usually do well in mathematics” or “Mathematics is harder for me than for many of my classmates” (Foy, Arora, & Stanco, 2013b, p. 11).
Mathematics assessment framework.

The assessment framework is broken down into two dimensions: content and cognitive. In mathematics, which shall be focused on for the remainder of this paper, the content dimension refers to the math skills or understanding needed to complete the tasks (e.g., multiplication, perimeter, etc.) (Mullis, Martin, Ruddock, O’Sullivan, and Preuschoff, 2009, p. 19). The cognitive dimension refers to the projected cognitive demand required of a student to answer a given mathematics question (p. 20). As is written within the assessment framework document, it is the “thinking process to be assessed (that is, knowing, applying, and reasoning). The cognitive domains describe the sets of behaviors expected of students as they engage with the mathematics content” (p. 20).

Both the content and cognitive dimensions are further divided into what is referred to as domains. The fourth grade domains are slightly different from the eighth domains because different curricula are covered at each of these grade levels. Table 3.3 provides information about each dimension and domain by grade level. The percentage of test items devoted to each domain is listed in parentheses. More details about the content and cognitive dimensions of the fourth and eighth grade can be found in Appendix B.
Table 3.3. TIMSS: A Comparison of Mathematics Dimensions and Domains of the Fourth and Eighth Grades.

<table>
<thead>
<tr>
<th>CONTENT DIMENSION</th>
<th>GRADE 4 DOMAIN:</th>
<th>GRADE 8 DOMAIN:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number (50%)</td>
<td>Number (30%)</td>
<td></td>
</tr>
<tr>
<td>Geometric Shapes and Measures (35%)</td>
<td>Geometry (20%)</td>
<td></td>
</tr>
<tr>
<td>Data Display (15%)</td>
<td>Data and Chance (20%)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>COGNITIVE DIMENSION</th>
<th>GRADE 4 DOMAIN:</th>
<th>GRADE 8 DOMAIN:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowing (40%)</td>
<td>Knowing (35%)</td>
<td></td>
</tr>
<tr>
<td>Applying (40%)</td>
<td>Applying (40%)</td>
<td></td>
</tr>
<tr>
<td>Reasoning (20%)</td>
<td>Reasoning (25%)</td>
<td></td>
</tr>
</tbody>
</table>

Notes.
1. Percentages relate to the percentage of test items devoted to that domain.
2. Adapted from Mullis, Martin, Ruddock, O'Sullivan, and Preuschoff, 2009, p. 20.
**Item types.**

In mathematics, students are tested with multiple-choice and constructed-response items to test content knowledge and ability to maneuver varying cognitive demands (Mullis & Martin, 2011d, p. 6-7). Students are offered 4 answer choices for multiple-choice items, where one answer is correct, referred to as the key, and the others are incorrect, referred to as distractors (p.11). About half of the questions on TIMSS 2011 were multiple-choice items (p.11). Each of these types of questions has been estimated to take students 1 minute to answer (p.7). Figure C.1 in Appendix C provides an example of a multiple-choice item.

Constructed-response items are estimated to take students about 1-3 minutes each. These kinds of questions differ from traditional multiple-choice questions in that they “allow students to provide explanations, support an answer with reasons or numerical evidence, draw diagrams, or display data” (Mullis & Martin, 2011d, p.7). Constructed-responses are worth 1 or 2 points, depending on the demand of the question (p. 18). For a 1 point item, 1 point is awarded if “the student has completed the task correctly,” and a 0 is given if the answer was “incorrect, irrelevant, or incoherent” (p. 18). For a 2 point test item, 2 points are awarded if the answer was “complete and correct,” 1 point is given if the answer is partially correct, and 0 points are given for an “incorrect, irrelevant, or incoherent” response (p. 18). Figures C.2 and C.3 in Appendix C provide examples of one and two point test items. Accompanying each constructed response item is the scoring guide used by scorers to award points.
All questions are reviewed for developmental appropriateness and to avoid biases such as putting one group of people to an unfair disadvantage over another (Mullis & Martin, 2011d, p. 8-9). As written in the TIMSS 2011 item writing guideline, “An international study requires special attention to the diversity of environments, backgrounds, beliefs, and more among students in participating countries” (p. 8).

**Item blocks.**

When students take the TIMSS mathematics assessments, not all students are given all of the same items. A number of test items are brought together in a block to create a unit of items. There are 10-14 test items included within each block in the fourth grade and 12-18 items in the eighth grade (Mullis, Martin, Ruddock, O’Sullivan, & Preuschoff, 2009, p. 123).

The fourth grade question presented in Figure C.1 of Appendix C appeared in test block labeled M04. For block M04, the Figure C.1 question would appear as well as 11-17 other questions with it. The number of questions and test questions of M04 are the same for each test booklet containing the M04 block. In the TIMSS 2007 and 2011 assessments, 14 blocks were created to test mathematics skills per student test booklet (p. 123). Some of these blocks were composed of new test items and some from years previous. The blocks of questions used in previous years were given to students in order to see how students’ abilities in answering those questions may have changed from one TIMSS year to another. For more information about the distribution of 2007 fourth grade and 2011 eighth grade TIMSS blocks, please see Appendix D.
Background questionnaires.

As previously mentioned, the National Research Coordinators (NRC) evaluate the assessment frameworks on which the test items are based. They are also responsible for reviewing the background questionnaires taken by administrators, teachers, and students of participating schools. These questionnaires ask questions about school resources (e.g., textbooks, technology, etc.), time on learning, education of parents, attitudes about subject matter, frequency of homework, number of books at home, etc. (Martin & Mullis, 2012, p. 2). In 2011, background questionnaires were given as part of TIMSS and the Progress in International Reading Literacy Study (PIRLS), international reading assessment, tests because they were both scheduled to be administered during the same year and (for many countries) to the same sample of students (Mullis & Martin, 2011g, p. 1). TIMSS and PIRLS are both projects of the International Study Center at Boston College.

In the fourth and eighth grades, student, teacher, and school questionnaires were given (Mullis & Martin, 2011g, p. 1). Student questionnaires asked students about their home backgrounds, attitudes toward school, and attitudes and behaviors towards particular subject matter (i.e., mathematics, science, and/or reading) (p.1). Teacher questionnaires asked teachers about their feelings regarding the school climate; preparedness to teach mathematics, reading, and/or science; their education; and their classroom coverage of subject matter tested by TIMSS (p. 1). In the fourth grade, teachers were asked questions about reading, mathematics, and/or science within the same booklet. However, in the eighth grade, teachers were given subject specific questionnaires (p. 2). Mathematics teachers completed mathematics teacher questionnaires (p. 2). School questionnaires asked school principals to respond to questions about student demographics, school resources, educational programs, and school climate (TIMSS 2011
context questionnaires, n.d., para. 5).

**The reporting and scaling of context questionnaires.**

The scales of the student context questionnaires test ask questions about general background information, such as parents’ education levels or attitudes about learning and instruction. Questions are asked of students in terms of their agreement with variety of statements (e.g., I usually do well in mathematics.). In most cases, students rate each statement by selecting either: agree a lot, agree a little, disagree a little, or disagree a lot (Martin, Mullis, Foy, and Arora, 2012, p.1). Using the Rasch partial credit model of Item Response Theory (IRT), each of these four responses are then typically categorized as a high, middle, or low value for the construct in TIMSS reports. That is, responses would then show if a student had high confidence in mathematics (labeled in a TIMSS report as “confident”), middle or moderate confidence in mathematics (labeled “somewhat confident”), or low confidence in mathematics (labeled “not confident”) (p.1).

Basically, if a student was to answer 5 out of 5 questions in the affirmative (as they positively related to confidence in mathematics), it is very likely that the student would identify him or herself as very confident in mathematics if asked explicitly about mathematics confidence (p.2). The likelihood of students identifying themselves with specific constructs based on their responses is how those low, middle, or high values are assigned to students (p.2).

Answers to questions are given numerical values. To a positive statement connected to confidence, such as, “Mathematics is easy for me,” would result in a score of 0 for “disagree a lot” to a score of 3 for “agree a lot.” For one scale, such as *Students Confident in Mathematics*, all the scores are summed for a raw score. The raw score is then converted to a scaled score, which relates to high, middle, or low values of that scale. Cut points are determined to provide a
reference for a scaled score as to the cut off scaled scores are for high, middle, or low values (p.5). Later, Cronbach’s Alpha is used to test the reliability scale. Pearson’s correlation is used to look at the relationship between a scale, such as Students Confident in Mathematics, and mathematics achievement (p.4).

**Achievement means.**

**Plausible values.**

Since not all students’ test booklets have the same test item blocks (e.g., only some test books will use test block M04), a plausible value is calculated to estimate students’ academic performances as if students were assigned all 14 blocks of test questions. Plausible values are assigned to students based on the results of their survey data and their response patterns to the questions they answered (Mullis & Martin, 2011f, p.4). They predict how a student would perform based on item and background questionnaire responses given “students with similar response patterns and background characteristics in the sampled population” (Mullis & Martin, 2011f, p. 6). They are not intended to obtain an individual’s score (p.6). In TIMSS reports, five plausible value estimates are generated for each student using the International Database (IDB) Analyzer (Version 3.0) with the assistance of SPSS. The variability between plausible values are then measured and each plausible value average is given a standard error to indicate the level of uncertainty of each achievement mean (plausible value) calculated for each student in a sample.

What’s beneficial to using this method is that it essentially assigns more questions to students without inundating students with questions, more test items can be analyzed as a result, and it decreases an amount of measurement error by calculated errors through 5 plausible values (Mullis & Martin, 2011f, p.4). Therefore, it is important to run TIMSS data using all 5 plausible values to obtain achievement means and standard errors than to run the data with just one.
plausible value set. To do the latter would increase the standard error of calculated achievement means.

**Mathematics international benchmarks.**

Once students’ plausible values are processed and achievement means identified, means are labeled with international benchmarks. These benchmarks describe the achievement of a country, of a gender, or other group of interest.

Means are labeled as: Advanced International Benchmark (625), High International Benchmark (550), Intermediate International Benchmark (475), or Low International Benchmark (400) (Mullis, 2012, p.3; Mullis, Erberer, & Preuschoff, 2008, p.339).

Items were anchored as:

1. Advanced International Benchmark (625) if at least 65% of the students performed in this range and less than 50% were at the High International Benchmark
2. High International Benchmark (550) if at least 65% scored within this range and less than 50% were at the Intermediate Benchmark.
3. Intermediate Benchmark (475) if at least 65% scored within this range and less than 50% were at the Low International Benchmark.
4. Low International Benchmark (400) if at least 65% scored within this range

(Mullis, 2012, p.3)
This Dissertation’s Methods

The objective of the first part of this chapter was to discuss the methods of the TIMSS assessment to help to create an understanding of the workings of the test including test design, sampling processes, and samples relevant to this dissertation. The second part of this chapter is devoted to the methodologies utilized in this dissertation to look at the relationships between gender, self-efficacy, and student achievement in the fourth grade TIMSS 2007 and the eighth grade TIMSS 2011.

The 2007-2011 cohort.

The importance of looking at the fourth grade and eighth grade students within this cohort is that the fourth graders of 2007 become the eighth graders of 2011. Though different students take the test in 2007 and 2011, this sample of students chosen to represent the population of fourth and eighth grade students of the United States may show how mathematics achievement and the effect of other variables on this achievement may have changed overtime.

Analyses.

Before analyzing the relationships of self-efficacy, gender, and mathematics achievement, the self-efficacy variables needed to be validated as a scale\(^6\).

---

\(^6\) Prior to the factor analysis, a correlation and regression analysis were conducted on mathematics means and self-efficacy, as a continuous variable, to explore whether or not there was a relationship between the two. For more information about this preliminary work, please refer to Appendix E.
To do this, I used the variables that overlapped from the Students Confident in Mathematics (SCM) scale from TIMSS 2007 and 2011. The SCM scale for fourth grade and eighth grade are presented in tables 3.4 and 3.5.

For items on this scale, students had to rate each statement with:

Agree a lot = 1
Agree a little = 2
Disagree a little = 3
Disagree a lot = 4
Table 3.4. TIMSS 2007: Students Confident in Mathematics Scale (SCM), Grade 4

Stem:
How much do you agree with these statements about learning mathematics?

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS4MAWEL</td>
<td>I usually do well in mathematics.</td>
</tr>
<tr>
<td>AS4MACLM</td>
<td>*Mathematics is harder for me than many of my classmates.</td>
</tr>
<tr>
<td>AS4MANOT</td>
<td>*I’m just not good at mathematics.</td>
</tr>
<tr>
<td>AS4MAQKY</td>
<td>I learn things quickly in mathematics.</td>
</tr>
</tbody>
</table>

* = reverse coded

Ratings:
Agree a lot = 1
Agree a little = 2
Disagree a little = 3
Disagree a lot = 4

Notes:
2. Negative statements (denoted by *) were reverse coded so that positive and negative statements wouldn’t negate each other.
Table 3.5. TIMSS 2011: Students Confident in Mathematics Scale (SCM), Grade 8

Stem: How much do you agree with these statements about learning mathematics?

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Statements:</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSBM16A</td>
<td>*I usually do well in mathematics.</td>
</tr>
<tr>
<td>BSBM16B</td>
<td>Mathematics is more difficult for me than many of</td>
</tr>
<tr>
<td></td>
<td>my classmates. (In the fourth grade, this statement</td>
</tr>
<tr>
<td></td>
<td>is slightly different. It reads, “Mathematics is</td>
</tr>
<tr>
<td></td>
<td>harder for me than many of my classmates.”</td>
</tr>
<tr>
<td></td>
<td>However, since the wording is so close, it was</td>
</tr>
<tr>
<td></td>
<td>still selected for analysis.)</td>
</tr>
<tr>
<td>BSBM16C</td>
<td>Mathematics is not one of my strengths.</td>
</tr>
<tr>
<td>BSBM16D</td>
<td>*I learn things quickly in mathematics.</td>
</tr>
<tr>
<td>BSBM16E</td>
<td>Mathematics makes me confused and nervous.</td>
</tr>
<tr>
<td>BSBM16F</td>
<td>*I am good at working out difficult mathematics</td>
</tr>
<tr>
<td></td>
<td>problems.</td>
</tr>
<tr>
<td>BSBM16G</td>
<td>*My teacher thinks I can do well in mathematics</td>
</tr>
<tr>
<td></td>
<td>&lt;programs/classes/lessons&gt; with difficult</td>
</tr>
<tr>
<td></td>
<td>materials.</td>
</tr>
<tr>
<td>BSBM16H</td>
<td>*My teacher tells me I am good at mathematics.</td>
</tr>
<tr>
<td>BSBM16I</td>
<td>Mathematics is harder for me than any other</td>
</tr>
<tr>
<td></td>
<td>subject.</td>
</tr>
</tbody>
</table>

* = reverse coded (Due to an error in the 2011 TIMSS database, positive statements had to be recoded instead of the negative.)

Ratings:
Agree a lot = 1
Agree a little = 2
Disagree a little = 3
Disagree a lot = 4

Notes:
1. Obtained from The TIMSS 2011 Students Confident in Mathematics Scale, eighth grade, 2012.
2. Positive statements (denoted by *) were reverse coded so that positive and negative statements wouldn’t negate each other. Usually, negative statements are reverse coded but due to a database error, the inverse needed to be completed for an accurate depiction of SCM results.
The variables available at both grade levels were:

1. I usually do well in mathematics. (AS4MAWEL/ BSBM16A)
2. Mathematics is harder for me than many of my classmates. (AS4MACLM)
   Mathematics is more difficult for me than many of my classmates. (BSBM16B)
3. I learn things quickly in mathematics. (AS4MAQKY/BSBM16D)

To prepare for analysis, the variables were recoded so that the lower the number, the higher the self-efficacy value (See the notes in tables 3.4 and 3.5.). After recoding and averaging values into a continues scale, the students’ relative self-efficacy ratings were categorized into three meaningful categories:

1 = high self-efficacy = average less than or equal to 2
2 = medium self-efficacy = average greater than 2 but less than 3
3 = low self-efficacy = average greater than or equal to 3

This categorization method is typically utilized in TIMSS reports. “Agree” ratings averaging less than or equal to 2 are used to indicate higher self-efficacy, and “disagree” ratings averaging greater than or equal to 3 are used to indicate lower self-efficacy. Ratings in between higher and lower self-efficacy levels have been identified as medium self-efficacy by values in between the high and low self-efficacy levels, averages greater than 2 but less than 3.

7 Though the wording of variables AS4MACLM and BSBM16B is different, during data analysis, I treated them as the same since the wording differed slightly.
An important distinction: Confidence versus self-efficacy.

Though the variables selected for analysis come from a scale called, “Students Confident in Mathematics Scale,” they are not confidence markers. They are statements of self-efficacy.

To revisit Bandura’s (1997) position on this, outlined in greater detail in chapter 2, confidence is a descriptor used to label overall feelings of self-worth. Confidence is not specific to a domain, in this case of this dissertation, the domain of mathematics. Self-efficacy on the other hand, is a term used to describe one’s perceived abilities within a certain domain. Self-efficacy is a phrase used to self-label one’s capabilities in reference to something as specific as mathematics or even a particular portion of that field, such as algebra or geometry.

Based on this definition, all of the variables used for this study are better labeled self-efficacy than confidence. As a result, the summary variable will be referred to as a self-efficacy index not a confidence index, as it is identified in the TIMSS User Guides of 2007 and 2011.

After the self-efficacy variables were determined, a principal component analysis using a varimax rotation was conducted to see how well each variable measured the construct, self-efficacy. Also, reliability was analyzed using Cronbach’s alpha to see how consistently the variables measured the underlying construct.

With the assistance of the IDB Analyzer SPSS plugin, mathematics achievement was compared by self-efficacy level, grade level, and gender. Statistical significances were tested by examination of standard errors. A regression analysis was also conducted to look at the impact of gender and self-efficacy as predictors of mathematics achievement.

In the next chapter, the results of these methods are presented.
Chapter 4: Results

Using the fourth grade TIMSS 2007 and the eighth grade TIMSS 2011 mathematics data, this dissertation examines the relationship of self-efficacy and achievement of boys and girls and how this relationship may have changed from the fourth grade 2007 TIMSS to the eighth grade 2011 TIMSS. In the first part of the data analysis, I investigated how well the variables I used measured self-efficacy using factor analysis. In the next step, I compared the self-efficacy of boys and girls within the 2007 and 2011 cohort to see if the gender difference is significant. Finally, I tested if there was a relationship between gender, self-efficacy and mathematics achievement by way of regression analysis.

This chapter presents the results of these analyses. Fourth grade data for 2007 is always presented first followed by 2011 eighth grade data.

Self-Efficacy Measurement

As outlined in Chapter 3, I used selected variables from the TIMSS Self-Confidence in Learning Mathematics (SCM) scales, which were identical or similar at both study years and grade levels, to measure self-efficacy. I conducted principal component analysis using a varimax rotation to test the degree to which these variables measured the same construct. Tables 4.1 and 4.2 show factor loadings for each component variable. Below each table, the percent of variance in the factor explained by the component variables (R-Square), an indicator of scale validity is displayed. Cronbach’s alpha is also provided, as an indicator of scale reliability (i.e., how consistently the component variables measure the underlying construct).
Table 4.1. Factor Analysis of Self-Efficacy Variables of TIMSS 2007, Grade 4

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Statement</th>
<th>Factor Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS4MAWEL</td>
<td>I usually do well in mathematics.</td>
<td>.826</td>
</tr>
<tr>
<td>AS4MACLM</td>
<td>Mathematics is harder for me than many of my classmates.</td>
<td>.749</td>
</tr>
<tr>
<td>AS4MAQKY</td>
<td>I learn things quickly in mathematics.</td>
<td>.800</td>
</tr>
</tbody>
</table>

Percent Variance Explained ($R^2$): 63%
Cronbach’s Alpha Reliability Coefficient: $r = .684$

Table 4.2. Factor Analysis of Self-Efficacy Variables of TIMSS 2011, Grade 8

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Statement</th>
<th>Factor Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSBM16A</td>
<td>I usually do well in mathematics.</td>
<td>.848</td>
</tr>
<tr>
<td>BSBM16B</td>
<td>Mathematics is more difficult for me than many of my classmates.</td>
<td>.786</td>
</tr>
<tr>
<td>BSBM16D</td>
<td>I learn things quickly in mathematics.</td>
<td>.862</td>
</tr>
</tbody>
</table>

Percent Variance Explained ($R^2$): 69%
Cronbach’s Alpha Reliability Coefficient: $r = .771$
Factor analysis results.

For the fourth grade, each of the component variables loaded moderately high on the underlying factor (loadings between .749 and .826), indicating that they measure the underlying construct relatively well. The percentage of variance in the factor explained by the component variables was 63%, which is respectably high for a three variable scale and Cronbach’s alpha was $r = .684$, providing further evidence that the three component variables are valid and three component variables work well as a unit.

At the eighth grade factor loadings were even higher (between .786 and .862), again supporting the hypothesis that all three variables measure the same construct. Together they explained 69% of the variance and Cronbach’s alpha was $r = .771$ providing strong evidence for a valid and reliable scale.

It can be concluded that both fourth grade and eighth grade analyses of scale dimensionality have shown that the three variables measure a common construct, which based on theoretical considerations is referred to as self-efficacy.
Self-efficacy ratings and mathematics achievement

Using the three component variables, parallel self-efficacy scales were computed at both grade levels, which would allow for cohort comparisons. First component variables were averaged to create a continuous scale. Reverse coding was then applied to some statements before averaging (See tables 3.4 and 3.5 in Chapter 3 for information about which statements had to be reverse coded.). That is, statements were calibrated so that responses to the highest self-efficacy levels were coded as 1 and with the lowest self-efficacy levels were coded as 4. For example, for the fourth grade statement, “Mathematics is harder for me than many of my classmates.” Students responded with either:

1 = Agree a lot
2 = Agree a little
3 = Disagree a little
4 = Disagree a lot

When reverse coding statements, a 1 became a 4 to reflect the lowest self-efficacy level. The 2 became a 3, the 3 became a 2, and the 4 became a 1. By doing this, the ratings for this variable could then be averaged with the fourth grade statements, “I usually do well in mathematics.” and “I learn things quickly in mathematics.” The latter statements were rated using the same rating system. However, they were not reverse coded\(^8\).

\(^8\) An important note, due to an error in the 2011 TIMSS database, positive statements for the eighth grade variables had to be recoded instead of negative statements.
In order to better understand quantitative differences in self-efficacy and the relationships it has with mathematics achievement and gender, three self-efficacy levels were created: high, medium, and low. It is the purpose of these three levels to meaningfully differentiate between students with different levels of self-efficacy.

1 = High = On average at least agrees to all three statements
2 = Medium = On average neither agrees or disagrees
3 = Low = On average at least disagrees to all three statements

Once the three self-efficacy levels were defined, mathematics achievement of the 2007-2011 cohort was tabulated by these levels. Table 4.3 displays the results and shows that the higher the self-efficacy, the higher mathematics achievement. The differences are statistically significant between all three self-efficacy levels at both grades.
Table 4.3. Mathematics Achievement by Self-Efficacy

<table>
<thead>
<tr>
<th></th>
<th>Self-Efficacy</th>
<th>Mathematics Mean (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 4</td>
<td>High</td>
<td>548.64 (2.49)*</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>499.70 (2.62)*</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>481.30 (3.13)*</td>
</tr>
<tr>
<td>Grade 8</td>
<td>High</td>
<td>533.96 (2.73)*</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>486.97 (2.52)*</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>460.81 (3.54)*</td>
</tr>
</tbody>
</table>

Note: * = Statistically significant
In tables 4.4 and 4.5, mathematics achievement by self-efficacy levels is reported by gender. Looking at boys and girls individually, higher self-efficacy is associated with higher mathematics achievement. This is true at both grade levels within the 2007-2011 cohort. However, with the addition of gender as a grouping factor, not all of the differences are statistically significant.

In both grade levels, there are more boys at the high self-efficacy level than girls. In the fourth grade, when boys’ self-efficacy is high, it is associated with significantly higher mathematics achievements than girls. At the eighth grade, although the gender difference for students with high self-efficacy still favors boys, their advantage in mathematics achievement is smaller and not statistically significant.

At the medium self-efficacy level, boys’ mathematics achievement is lower than girls. This gender difference, however, is not statistically significant in the fourth grade but is in the eighth grade. Boys also perform worse than girls at the low self-efficacy levels, but at both these levels, the differences are not statically significant.

In sum, boys with high self-efficacy are at an advantage over girls in their mathematics achievement. This is not true for boys with medium or low self-efficacy. The analysis even shows that if self-efficacy is lower, boys could be at a disadvantage.

Table 4.5 summarizes the relationships described in tables 4.3 and 4.4. Statistical significances are also reported as described above. The relationships are further illustrated by line graphs in figures 4.1 and 4.2. The table and graphs show that the higher the self-efficacy, the higher mathematics achievement. However, overall, boys who have high self-efficacy perform better than girls of comparable self-efficacy, a difference that is statistically significant at the
fourth grade. At the low and medium self-efficacy levels, girls perform better than boys. However, the difference is not always statistically significant.
Table 4.4. TIMSS 2007: Fourth Grade Mathematics Achievement by Self-Efficacy and Gender

| Self-Efficacy | Boys | | | | Girls | | | |
|---------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|               | N               | %               | Mathematics Means (s.e) | Standard Deviation (s.e) | N               | %               | Mathematics Means (s.e) | Standard Deviation (s.e) |
| High          | 2643            | 70.07 %         | 551.81 (2.7)           | 72.78 (1.46)            | 2543            | 64.38 %         | 545.35 (2.85)           | 71.17 (1.41)            |
| Medium        | 772             | 20.83 %         | 497.38 (2.94)          | 65.49 (1.85)            | 909             | 22.45 %         | 501.76 (3.52)          | 64.98 (2.20)            |
| Low           | 333             | 9.10 %          | 480.74 (5.14)          | 60.73 (2.88)            | 493             | 13.17 %         | 481.67 (4.08)          | 62.41 (2.81)            |

Table 4.5. TIMSS 2011: Eighth Grade Mathematics Achievement by Self-Efficacy and Gender

| Self-Efficacy | Boys | | | | Girls | | | |
|---------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|               | N               | %               | Mathematics Means (s.e) | Standard Deviation (s.e) | N               | %               | Mathematics Means (s.e) | Standard Deviation (s.e) |
| High          | 3124            | 61.56 %         | 536.37 (2.92)           | 72.28 (1.59)           | 3038            | 57.46 %         | 531.49 (3.05)           | 72.16 (1.73)           |
| Medium        | 1287            | 25.04 %         | 483.09 (3.08)          | 68.35 (2.16)           | 1250            | 23.96 %         | 490.85 (2.76)          | 68.93 (2.13)           |
| Low           | 681             | 13.41 %         | 459.99 (3.88)          | 67.67 (2.70)           | 942             | 18.58 %         | 461.37 (4.99)          | 64.02 (2.99)           |

Table 4.6. Comparison of Achievement Means by Self-Efficacy and Gender

<table>
<thead>
<tr>
<th>Grade</th>
<th>High Self-Efficacy</th>
<th>Medium Self-Efficacy</th>
<th>Low Self-Efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>B&gt;G *</td>
<td>B&lt;G</td>
<td>B&lt;G</td>
</tr>
<tr>
<td>8</td>
<td>B&gt;G</td>
<td>B&lt;G *</td>
<td>B&lt;G</td>
</tr>
</tbody>
</table>

Notes. 1. B = boys’ mathematics achievement  
G = girls’ mathematics achievement  
2. Starred items are statistically significant.
Figure 4.1. TIMSS 2007: Fourth Grade Self-Efficacy Ratings and Mathematics Means by Gender

Note: Mathematics means are reported here without standard error calculations. Please refer to Table 4.4 for standard error information.

Figure 4.2. TIMSS 2011: Eighth Grade Self-Efficacy Ratings and Mathematics Means by Gender

Note: Mathematics means are reported here without standard error calculations. Please refer to Table 4.5 for standard error information.
Predicting Mathematics Achievement

In the following regression analysis, I looked at how well gender and self-efficacy predict the mathematics achievement for students in the 2007-2011 cohort.

The hypotheses tested were:

A. That gender is a significant predictor of mathematics achievement.

B. That self-efficacy is a significant predictor of mathematics achievement.

C. That the interaction between gender and self-efficacy does contribute to significantly predicting mathematics achievement.

The latter hypothesis was introduced because previous analysis had shown that boys at the high self-efficacy level have an advantage over girls, which however, was not consistent across self-efficacy levels. It diminishes with decreasing self-efficacy.

In all regression models, boys are coded 1 and girls are coded 0. Also, for the regression analyses two dummy coded self-efficacy variables were created: 1) High self-efficacy (=1) vs. medium and low self-efficacy (=0) and from now on referred to as high self-efficacy effect, and 2) high and medium self-efficacy (=1) vs. low self-efficacy (=0) and from now on referred to as medium self-efficacy effect. This coding scheme allowed looking at the effects associated with self-efficacy at both ends of the continuum.

To test the interaction between gender and self-efficacy, gender and self-efficacy levels were also multiplied to compute interaction terms.

Two interaction terms were created: 1) the interaction between gender (being a boy instead of a girl) and the high self-efficacy effect and 2) the interaction between gender (being a boy instead of a girl) and the medium self-efficacy effect.
Three distinct regression models were tested. Model 1 only looks at the effect gender has on mathematics achievement. In Model 2, the effect self-efficacy has on mathematics achievement after controlling for gender is tested. Finally, in Model 3, the effect of the interaction between gender and self-efficacy on mathematics achievement was tested. It was the purpose of this model to see if after controlling for gender and self-efficacy individually the association between the two variables still has a significant impact on how students perform.

Tables 4.6 and 4.7 display the regression results. Fourth grade data is displayed in the first table and eighth grade in the second.
Table 4.7. TIMSS 2007 Grade 4 Regression Models: Relationships between Mathematics Achievement, Gender, and Self-Efficacy

<table>
<thead>
<tr>
<th>Grade 4</th>
<th>Model 1: Gender $R^2 = .00$</th>
<th>Model 2: Gender and Self-Efficacy $R^2 = .13$</th>
<th>Model 3: Gender, Self-Efficacy and Gender-Self Efficacy Interaction $R^2 = .13$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics Achievement (Constant)</td>
<td>526.08 (2.75)</td>
<td>497.98 (3.38)</td>
<td>481.67 (4.08)</td>
</tr>
</tbody>
</table>

1. Gender | 5.98 (2.42) | 3.31 (2.25) | -0.94 (6.70) |
2. High Self-Efficacy | | 48.8 (2.32) | 43.59 (3.51) |
3. Medium Self-Efficacy | | 18.17 (3.40) | 20.09 (4.64) |
4. Interaction: Gender $\times$ High Self-Efficacy | | | 10.85 (4.45) |
5. Interaction: Gender $\times$ Medium Self-Efficacy | | | -3.45 (7.28) |

Note: Standard errors are reported in parentheses.

Table 4.8. TIMSS 2011 Grade 8 Regression Models: Relationships between Mathematics Achievement, Gender, and Self-Efficacy

<table>
<thead>
<tr>
<th>Grade 8</th>
<th>Model 1: Gender $R^2 = .00$</th>
<th>Model 2: Gender and Self-Efficacy $R^2 = .15$</th>
<th>Model 3: Gender, Self-Efficacy and Gender-Self Efficacy Interaction $R^2 = .15$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics Achievement (Constant)</td>
<td>507.75 (2.94)</td>
<td>460.49 (3.81)</td>
<td>461.37 (4.99)</td>
</tr>
</tbody>
</table>

1. Gender | 3.54 (2.25) | .78 (2.17) | -1.38 (5.83) |
2. High Self-Efficacy | | 46.98 (2.39) | 40.64 (2.73) |
3. Medium Self-Efficacy | | 26.09 (2.89) | 29.48 (4.81) |
4. Interaction: Gender $\times$ High Self-Efficacy | | | 12.63 (3.17) |
5. Interaction: Gender $\times$ Medium Self-Efficacy | | | -6.38 (6.63) |

Note: Standard errors are reported in parentheses.
Fourth grade.

**Model 1.** Regression equation: 526.08 + (5.98 × gender)

At the fourth grade level, gender on its own does not explain any variance in mathematics achievement ($R^2 = .00$). Boys do have a small advantage over girls in their mathematics achievement (approximately 6 points), but this difference is not statistically significant.

**Model 2.** Regression equation: 497.98 + (3.31 × gender) + (48.8 × high self-efficacy) + (18.17 × medium self-efficacy)

Self-efficacy explains 13% of the variance in mathematics achievement after controlling for gender, especially for boys being at the high self-efficacy level. At this level, boys have a substantial, almost 50-point, advantage over girls on the TIMSS 2007 mathematics assessment. This effect can also be found at the medium self-efficacy level. However, it is less than half of the size.

**Model 3.** Regression equation: 481.67 + (-.94 × gender) + (43.59 × high self-efficacy) + (20.09 × high self-efficacy) + (10.85 × interaction of gender × high self-efficacy) + (-3.45 × interaction of gender × medium self-efficacy)

Again, the high self-efficacy variables contribute the most to a male advantage. Beyond self-efficacy on its own, being a boy and having high self-efficacy is associated with an additional
11 percent advantage in mathematics achievement. This effect cannot be found at the medium self-efficacy level. Like model 2, 13% of the variance is explained by this model.

**Eighth grade.**

*Model 1.* Regression equation: 507.75 + (3.54 × gender)

Similarly to the fourth grade, gender on its own does not explain any variance in mathematics achievement ($R^2 = .00$).

*Model 2.* Regression equation: 460.49 + (.78 × gender) + (46.98 × high self-efficacy) + (26.09 × medium self-efficacy)

At the eighth grade (Table 4.7), 15% of the variance in mathematics achievement can be explained by this model. After controlling for gender, boys at the high self-efficacy level have an approximately 47-point advantage over girls in mathematics achievement. At the medium self-efficacy level this effect is only about half the size.

*Model 3.* Regression equation: 461.37 + (-1.38 × gender) + (40.64 × high self-efficacy) + (29.48 × medium self-efficacy) + (12.63 × interaction of gender × high self-efficacy) +(-6.38 × interaction of gender × medium self-efficacy)
Model 3 shows that beyond gender and self-efficacy on its own, the interaction between high self-efficacy and gender again is a significant predictor of mathematics achievement, an effect that does not pertain for the medium self-efficacy level. Again, mirroring model 2’s variance, 15% of the variance in mathematics achievement is explained by this model 3.

**Summary of Findings**

This chapter examined the relationship between gender, self-efficacy and mathematics achievements of students in the 2007-2011 cohort in the United States. The main findings were:

1. Gender on its own is not a significant predictor of mathematics achievement.
2. At both grades, boys at the high self-efficacy level have a significant advantage over girls in mathematics achievement. This effect remains as students move up thru the grade levels.
3. The findings are less clear at the low and medium end of the self-efficacy continuum.
Chapter 5: Conclusions

Chapter 5 offers an overview of this study’s investigation of the relationships between the mathematics achievements and self-efficacy of boys and girls of the United States who took the 2007 TIMSS assessment in the fourth grade or the 2011 TIMSS assessment in the eighth grade. The TIMSS data utilized was examined through 1) a factor analysis of the self-efficacy variables, 2) a comparison of mathematics achievement, self-efficacy, and gender, and 3) a regression analysis used to examine how well gender and self-efficacy can predict mathematics achievement. The findings are summarized within this chapter and implications, study limitations, and suggestions for future research are discussed.

Overview of the Study

Larry Summers, former Harvard President, made headlines when stating that boys are represented in greater numbers than girls in Science, Technology, Engineering, and Mathematics (STEM) fields in the United States due to a biological advantage in their mathematics-based aptitudes (Summers, 2005). In the United States, in particular, in the 2007 and 2011 cohort of students studied for this dissertation work, boys did outperform girls in mathematics (Mullis, Martin, & Foy [with Olson, Preuschoff, Erberber, Arora, & Galia], 2008; Mullis, Martin, Foy, & Arora, 2012;). However, in order for a biological advantage in mathematics to be a practical argument for the disparity in mathematics achievement between males and females, it would be reasonable to assume this type of mathematics disparity would exist across nations. According to recent PISA and TIMSS international results, this is not the case (Mullis, Martin, & Foy [with Olson, Preuschoff, Erberber, Arora, & Galia], 2008; Mullis, Martin, Foy, & Arora, 2012; OECD,
There are countries in which boys outperform girls in mathematics, girls outperform boys in mathematics, and countries in which boys and girls perform comparably, that is, with no statistically significant differences between the two groups.

With this in mind, I focused my work instead on socio-emotional factors to explain why in the United States boys in the 2007-2011 cohort outperformed girls. In particular, I looked at how gender, as a socialized construct, and self-efficacy may have contributed to this phenomenon.

Societal influences of gender in which what it means to be a boy or a girl has been cited to shape the interactions, philosophies, values, and self-belief systems of an individual (e.g., Davies, 1989; Foucault, 1980; Halpern, 2012; Helwig, Anderson, & Tindal, 2001; Tiedemann, 2000). These influences have been noted to shape likes and dislikes, which may have implications as to why males are more apt to choose STEM positions, and have been shown to shape academic achievement or underachievement (e.g., Beilock, Gunderson, Ramirez, & Levine, 2010 and Tiedemann, 2000).

Self-efficacy, the belief that a people have in terms of whether or not they have the capability to achieve a goal or accomplish a task, has been shown by researchers to act as a source of empowerment that may help people persevere through obstacles and help them to be successful (Bandura, 1993; Bandura, 1997; Fast, Lewis, Bryant, Bocian, Cardullo, Rettig, & Hammond, 2010; Pajares & Urdan, 2006). Self-efficacy has also proven to be a better predictor of mathematics achievement than past mathematics performances on tests (Pajares and Miller, 1994).
Findings of the Study

The primary aim of this dissertation was to provide information about gender and self-efficacy in relation to the mathematics achievements of boys and girls of the 2007 and 2011 cohort of the United States. The goal was to contribute to the discussion of whether or not socio-emotional factors such as gender and self-efficacy may explain the performances of boys in girls in mathematics as an alternative to biological explanations.

The study began by examining the self-efficacy variables that would be used for this study using a factor analysis. Variables that were used for this analysis were similar or identical on the 2007 fourth grade and 2011 eighth grade student background questionnaires. The results showed that the variables supported the same construct, which for this study was self-efficacy.

Next, self-efficacy responses from students were defined as high, medium, and low and mathematics achievement by self-efficacy level showed that the higher the self efficacy, the higher the achievement. This was true for both grade levels supports Bandura’s Theory of Self-Efficacy. In terms of mathematics, that is, the greater people believe that they are good at mathematics, the more likely they will achieve.

Then, mathematics achievements of boys and girls in the 2007-2011 cohort showed that when boys believed in their mathematics abilities in the high self-efficacy category, they would outperform girls. Though this was only statistically significant in the fourth grade, this finding was also present in the eighth grade.

In the medium and low self-efficacy categories, girls’ mathematics achievements were higher than boys. Though only statistically significant in the eighth grade for medium self-efficacy, studying the mean differences of the 2007 fourth grade and 2011 eighth grade levels
suggests that boys who have high levels of self-efficacy perform to higher degrees than girls. But, when boys’ self-efficacy levels are not high, they don’t perform as well as the girls. This does not seem to contradict Bandura’s theory that the higher the self-efficacy, the higher the mathematics achievement because boys and girls both have higher mean scores as self-efficacy levels increase.

Finally, the predictability of gender and self-efficacy to mathematics achievement on the fourth grade 2007 TIMSS and eighth grade 2011 TIMSS assessments were examined through a regression analysis. This was conducted to determine whether or not (1) gender was a significant predictor of mathematics achievement; (2) self-efficacy was a significant predictor of mathematics achievement; and (3) the interactions of gender and self-efficacy has an effect on mathematics achievement.

The results showed that gender alone is not a significant predictor of mathematics achievement. When accounting for self-efficacy and the interactions of self-efficacy and gender, gender has a decreasing effect on mathematics. So much so, that when accounting for gender, self-efficacy, and their interactions, gender puts boys at a disadvantage in mathematics achievement.

In terms of self-efficacy, high self-efficacy acts as the greatest contributor to boys having an advantage over girls in the fourth and eighth grade levels. This reflects the findings found in the comparisons of gender, self-efficacy levels, and mathematics achievements of boys and girls of the 2007-2011 cohort. The results from the regression analysis also support that mathematics achievement does not necessarily rise proportionally for boys and girls at the low, medium, and high self-efficacy levels.
Implications of the Study

This study has provided evidence that self-efficacy is a predictor of mathematics achievement and challenges the idea that boys have higher aptitudes in mathematics by biological design. In fact, it is hard to believe that one gender may be inferior to another in mathematics based on biology when international data does not support that one gender routinely outperforms another in mathematics.

As argued by civil rights leader, Robert Moses, “Mathematics education is a civil rights issue...In today’s world, economic access and full citizenship depend crucially on math and science literacy” (2001, p.5). With females currently being underrepresented in STEM positions, this is an issue (U.S. Department of Commerce, Economics and Statistics Administration, 2013). However, whether or not this gender disparity within the STEM field has to do with how well students perform in mathematics is not well defined. In fact, as previously discussed, the US Department of Commerce has found reasons outside of a lack of mathematical ability for females to choose non-STEM positions. These include difficulty balancing childcare and careers, few female role models in STEM, and strong gender stereotypes discouraging women to pursue STEM degrees and jobs (Beede, Julian, Langdon, McKittrick, Khan, & Doms. 2011, p. 8).

In thinking about strong gender stereotypes, it’s important for schools, families, and communities to recognize how inequities in the STEM field or in mathematics performances can be contributed by biased discursive practices that may dissuade females from pursing engineering jobs that are stereotypically male positions (Else-Quest, Hyde, Lynn, 2010; Guiso, Monte, Sapienza, & Zingales, 2008; Riegle-Crumb, 2005) by facilitating negative environments that may negatively impact self-efficacy, academic achievement, and educational opportunities (Bandura, 1980).
Morally and ethically, schools must be aware of the discursive practices and its implications within the classroom in order to protect students from biases that may negatively influence student self-efficacy. As Foucault (1980) expressed, schools have power to reproduce and shape norms. The activities that take place within schools have the ability to change discourse and thereby change students’ beliefs. Self-beliefs are critical to students’ academic success. With that said, it should also be recognized that beliefs and bias are not just consciously transmitted but subconsciously as well (Helwig et al., 2001). If teachers harbor beliefs that one group is better than the other in the mathematics, even if those sentiments are never verbalized in class, it may still be demonstrated. For example, teachers’ biases and beliefs may be implied through their level of question difficulty. If a group of students is routinely asked more difficult or probing questions than another group, it may be perceived that the former group is viewed by the teacher as more capable. This may effect the self-efficacy level of students. It may increase self-efficacy of the perceived more capable students and reduce the self-efficacy of the perceived less capable students. With this being said, certainly there should still be efforts made to demonstrate that a teacher believes in all students’ capabilities to be skilled at mathematics. However, other behaviors and discourse patterns that may expose biases tacitly also need to be reflected on as much as is consciously possible.
Limitations of the Study

As previously described, participants of TIMSS are selected to represent the demographics of their country. Participants were selected based on geographic location, socioeconomic status, school type, and degree of urbanization, and minority/ethnicity status. Additionally, large sample sizes are used in the fourth and eighth grade. Furthermore, field testing of potential test questions and reliability and validity tests are also carried out to make sure that the test measures what was intended to be measured. The TIMSS assessment requires so much time and care in order to be considered ready for administration that it takes years to complete. One TIMSS assessment development to implementation cycle takes 4 years.

However, even with all the checks and balances in place, there are limitations to the data used in this study. First, TIMSS data is not a randomized experiment. Additionally, this data cannot be used to describe causal relationships. This study should also not be used to generalize to other countries. This cohort study only examines the relationships of students who took the fourth grade 2007 or the eighth grade 2011 TIMSS mathematics assessments of the United States. On a similar note, the students used in this study were from the same population but not the same sample. The fourth grade students of 2007 and the eighth grade students of 2011 did not take both assessments. TIMSS is not designed to be a longitudinal test. Lastly, plausible values were used to determine mathematics achievement. Though the use of plausible values is highly reliable, it is not as reliable as data that comes from students taking an entire test. Finally, the wording of one of the background questionnaires statements was slightly different in the fourth grade than in the eighth grade. In the fourth grade, the statement reads, “Mathematics is harder for me than many of my classmates.” In the eighth grade, it’s “Mathematics is more difficult for me than many of my classmates.” The slight difference in wording may have affected the self-efficacy outcome.
Despite these limitations, the results presented in this dissertation were consistent with research presented in the literature.

**Suggestions for Future Research**

It would be valuable to conduct a study where students take the fourth grade TIMSS test and four years later the same students take the eighth grade TIMSS test. Having the opportunity to work on a longitudinal study like this would allow for an examination of changes in self-efficacy, if any, and its effect on mathematics achievement.

Another possible avenue for future research might be to repeat this cohort study in the United States for other TIMSS tests in order to see if the findings in this dissertation are unique to the 2007-2011 cohort or may be found in other cohorts of U.S. students. Additionally, it would be interesting to study the self-efficacy levels of boys and girls and their mathematics achievement in countries in which boys outperform girls, girls outperform boys, and boys and girls perform commensurately in mathematics in order to see if any trends emerge within or between these groups.

Lastly, it would be fascinating to look at the results of a study in which gender, self-efficacy, and reading achievement are examined using Progress in International Reading Literacy Study (PIRLS) data. PIRLS is a reading assessment produced by the same organization that produces the TIMSS. In a study like this, it would be interesting to see if gender alone continues to be a poor predictor of achievement, in this case, reading achievement, and high self-efficacy a great predictor.
References


Appendices
## Appendix A: TIMSS Participating Countries and Benchmarking Participants

*Table A.1. TIMSS 2007: List of Participating Countries*

<table>
<thead>
<tr>
<th>Participating Countries</th>
<th>Benchmarking Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td>Slovenia</td>
</tr>
<tr>
<td>Armenia</td>
<td>Sweden</td>
</tr>
<tr>
<td>Australia</td>
<td>Syrian Arab Republic</td>
</tr>
<tr>
<td>Austria</td>
<td>Thailand</td>
</tr>
<tr>
<td>Bahrain</td>
<td>Tunisia</td>
</tr>
<tr>
<td>Bosnia and Herzegovina</td>
<td>Ukraine</td>
</tr>
<tr>
<td>Botswana</td>
<td>United States</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>Yemen</td>
</tr>
<tr>
<td>Chinese Taipei</td>
<td></td>
</tr>
<tr>
<td>Columbia</td>
<td></td>
</tr>
<tr>
<td>Cyprus</td>
<td></td>
</tr>
<tr>
<td>Czech Republic</td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td></td>
</tr>
<tr>
<td>Egypt</td>
<td></td>
</tr>
<tr>
<td>El Salvador</td>
<td></td>
</tr>
<tr>
<td>England</td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td></td>
</tr>
<tr>
<td>Ghana</td>
<td></td>
</tr>
<tr>
<td>Honduras</td>
<td></td>
</tr>
<tr>
<td>Hong Kong SAR</td>
<td></td>
</tr>
<tr>
<td>Hungary</td>
<td></td>
</tr>
<tr>
<td>Indonesia</td>
<td></td>
</tr>
<tr>
<td>Iran, Islamic Rep. of</td>
<td></td>
</tr>
<tr>
<td>Israel</td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td></td>
</tr>
<tr>
<td>Jordan</td>
<td></td>
</tr>
<tr>
<td>Kazakhstan</td>
<td></td>
</tr>
<tr>
<td>Korea, Rep. of</td>
<td></td>
</tr>
<tr>
<td>Kuwait</td>
<td></td>
</tr>
<tr>
<td>Latvia</td>
<td></td>
</tr>
<tr>
<td>Lebanon</td>
<td></td>
</tr>
<tr>
<td>Lithuania</td>
<td></td>
</tr>
<tr>
<td>Malaysia</td>
<td></td>
</tr>
<tr>
<td>Malta</td>
<td></td>
</tr>
<tr>
<td>Mongolia</td>
<td></td>
</tr>
<tr>
<td>Morocco</td>
<td></td>
</tr>
<tr>
<td>The Netherlands</td>
<td></td>
</tr>
<tr>
<td>New Zealand</td>
<td></td>
</tr>
<tr>
<td>Norway</td>
<td></td>
</tr>
<tr>
<td>Oman</td>
<td></td>
</tr>
<tr>
<td>Palestinian Nat’l Auth.</td>
<td></td>
</tr>
<tr>
<td>Qatar</td>
<td></td>
</tr>
<tr>
<td>Romania</td>
<td></td>
</tr>
<tr>
<td>Russian Federation</td>
<td></td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td></td>
</tr>
<tr>
<td>Scotland</td>
<td></td>
</tr>
<tr>
<td>Serbia</td>
<td></td>
</tr>
<tr>
<td>Singapore</td>
<td></td>
</tr>
<tr>
<td>Slovak Republic</td>
<td></td>
</tr>
</tbody>
</table>

Note. Obtained from Mullis & Martin, 2008.
Table A.2. TIMSS 2011: List of Participating Countries

<table>
<thead>
<tr>
<th>Country</th>
<th>Country</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Armenia</td>
<td>Kazakhstan</td>
<td>Sweden</td>
</tr>
<tr>
<td>Australia</td>
<td>Korea, Rep. of</td>
<td>Syrian Arab Republic</td>
</tr>
<tr>
<td>Austria</td>
<td>Kuwait</td>
<td>Thailand</td>
</tr>
<tr>
<td>Azerbaijan</td>
<td>Lebanon</td>
<td>Tunisia</td>
</tr>
<tr>
<td>Bahrain</td>
<td>Lithuania</td>
<td>Turkey</td>
</tr>
<tr>
<td>Belgium (Flemish)</td>
<td>Macedonia</td>
<td>Ukraine</td>
</tr>
<tr>
<td>Botswana</td>
<td>Malaysia</td>
<td>United Arab Emirates</td>
</tr>
<tr>
<td>Chile</td>
<td>Malta</td>
<td>United States</td>
</tr>
<tr>
<td>Chinese Taipei</td>
<td>Morocco</td>
<td>Yemen</td>
</tr>
<tr>
<td>Croatia</td>
<td>The Netherlands</td>
<td></td>
</tr>
<tr>
<td>Czech Republic</td>
<td>New Zealand</td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>Northern Ireland</td>
<td></td>
</tr>
<tr>
<td>England</td>
<td>Norway</td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>Oman</td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td>Palestinian Nat’l Auth.</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>Poland</td>
<td></td>
</tr>
<tr>
<td>Ghana</td>
<td>Portugal</td>
<td></td>
</tr>
<tr>
<td>Honduras</td>
<td>Qatar</td>
<td></td>
</tr>
<tr>
<td>Hong Kong SAR</td>
<td>Romania</td>
<td></td>
</tr>
<tr>
<td>Hungary</td>
<td>Russian Federation</td>
<td></td>
</tr>
<tr>
<td>Indonesia</td>
<td>Saudi Arabia</td>
<td></td>
</tr>
<tr>
<td>Iran, Islamic Rep. of</td>
<td>Serbia</td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td>Singapore</td>
<td></td>
</tr>
<tr>
<td>Israel</td>
<td>Slovak Republic</td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>Slovenia</td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>South Africa</td>
<td></td>
</tr>
<tr>
<td>Jordan</td>
<td>Spain</td>
<td></td>
</tr>
</tbody>
</table>

Benchmarking Participants
Abu Dhabi, UAE
Alabama, USA
Alberta, Canada
California, USA
Colorado, USA
Connecticut, USA
Dubai, UAE
Florida, USA
Indiana, USA
Massachusetts, USA
Minnesota, USA
North Carolina, USA
Ontario, Canada
Quebec, Canada

Note. Obtained from Mullis & Martin, 2011b.
Appendix B: TIMSS Content and Cognitive Dimensions

Content Dimension

In the fourth grade, students are tested in three content areas: number, geometric shapes and measures, and data display. What follows is a brief overview of each domain.

Number is examined through four topics: (1) whole number; (2) fractions and decimals; (3) number sentences with whole numbers; and (4) patterns and relationships (Mullis, Martin, Ruddock, O’Sullivan, and Preuschoff, 2009, p.23). It is believed that:

[a]t the fourth grade, students should have developed number sense and computational fluency, understand the meanings of operations and how they relate to one another, and be able to use numbers and operations (i.e., add, subtract, multiply, and divide) to solve problems. They should be familiar with a range of number patterns, exploring the relationships between the numbers which are in a pattern or are used to derive it.

(p. 22-23)

The Geometric Shapes and Measures domain is measured through two areas: (1) points, lines, and angels, and (2) two- and three-dimensional shapes (p.26). Spatial sense is examined through the assessment of students’ abilities to:

identify and analyze the properties and characteristics of lines, angles, and a variety of geometric figures, including two- and three-dimensional shapes, and to provide explanations based on geometric relationships. This domain includes understanding informal coordinate systems and using spatial visualization skills to relate between two- and three-dimensional representations of the same shape.

(p. 26)
Through Data Display questions, students are primarily asked to (1) read and interpret data as well as (2) organize and represent data (Mullis, Martin, Ruddock, O’Sullivan, and Preuschoff, 2009, p.28). Reading and interpreting data tasks include the demonstration of table, pictograph, bar graph, and pie chart knowledge (p. 28). Not only do students have to read these types of data representations, they also have to be able to use data that has been gathered together and accurately create a data display fits a given context or test directive (p.28).

In the eighth grade, students are tested in number, geometry, data and chance, and algebra. The difficulty level of these questions make these test items standout from fourth grade items, even if within the actual assessment framework objectives the wording appears similar to that in the fourth grade (p. 29)

Topics in number include (1) whole numbers; (2) fractions and decimals; (3) integers; and (4) ratios, proportions, and percents (p. 30).

Computational work in the eighth grade is focused on fractions and decimals not whole numbers (p. 30). Problems are set in real world as well as purely mathematical contexts (p. 30). Students are also required to work flexibly; to convert between fractions, decimals, and percentages; to find equivalent ratios; and to understand the order and magnitude of integers (p. 31).

Geometry is concerned with: (1) geometric shapes, (2) geometric measurement, and (3) location and movement (Mullis et al., 2009, p. 34).

Like the fourth grade, this topic has been designed to test spatial sense. However, unlike the fourth grade:

The cognitive range extends from making drawings and constructions to mathematical reasoning about combinations of shapes and transformation. Students will be asked to
describe, visualize, draw, and construct a variety of geometric figures, including angles, lines, triangles, quadrilaterals, and other polygons. Students should be able to combine, decompose, and analyze compound shapes. (p. 34)

Students should also be able to navigate the Cartesian plane; apply perimeter, area, circumference, and volume formulas; and be able to draw, measure, and estimate give angles and lines (p. 36).

Data and Chance includes the topics of (1) data organization and representation; (2) data interpretation; and (3) chance (p. 36). Similarly to the fourth grade, students are asked to organize and interpret data. However, in the eighth grade, students are also required to know how to read and correctly apply continuous data using line graphs and calculate means, medians, modes, ranges, and describe the shape of the data spread “in general terms” (p. 38).

Of chance questions, students are tested on how well they apply the terms: certain, more likely, equally likely, less likely, or impossible to a given situation (Mullis et al., 2009, p. 38). They must also estimate the chances of future outcomes in order to solve problems (p. 38).

Within algebra, students are assessed on: (1) patterns, (2) algebraic expressions, and (3) equations/formulas and functions (p. 32). “The algebraic content domain includes recognizing and extending patterns, using algebraic symbols to represent mathematical situations, and developing fluency in producing equivalent expressions and solving linear equations” (p. 32).

These algebraic skills encompass extension of numeric and symbolic patterns, the modeling of situations through the development of expressions, and solving simultaneous equations (p. 33).
Cognitive Dimension

The domains under this dimension are the same for the fourth and eighth grade. These domains are: knowing, applying, and reasoning.

Knowing requires recalling facts and concepts. One premise for including factual questions is that “[f]acts encompass the factual knowledge that provides the basic language of mathematics, and the essential mathematical facts and properties that form the foundation of mathematical thought (p. 41). Basic mathematical understanding is crucial to higher order mathematical thinking and execution. Key terms used to help define this domain are: recall, recognize, compute, retrieve, measure, and classify/order (p. 42).

Application of mathematical ideas and concepts to given situations is “fundamental to success in the subject” (p. 43). This category of problems most imitates those found in school textbooks (p. 43). Key terms used when developing these questions include select (as in select a method), represent (relating to data representations), model, and implement (as in drawing shapes using specific parameters), and solve routine problems (p. 44). Routine problems labels problems similar to those used in classrooms (p. 44).

Within the reasoning domain, students are asked to solve non-routine problems to situations less likely to be encountered in school than routine problems (p. 45). “The third domain, reasoning, goes beyond the solution of routine problems to encompass unfamiliar situations, complex contexts, and multi-step problems” (p. 40). It requires students to use logic and reasoning to solve problems. Key terms within this domain include: analyze, generalize/specialize, integrate/synthesize, justify, and solve non-routine problems (p. 46).
Appendix C: TIMSS Sample Test Items

Figure C.1. TIMSS: Multiple-Choice Item, Fourth Grade

Which group of numbers is ordered from LARGEST to SMALLEST?

A  10,011; 10,110; 11,001; 11,100
B  10,110; 10,011; 11,100; 11,001
C  11,001; 11,100; 10,110; 10,011
D  11,100; 11,001; 10,110; 10,011

Note. Obtained from Mullis & Martin 2011d.
Figure C.2. TIMSS: One Point Constructed Response, Eighth Grade

Joe knows that a pen costs 1 zed more than a pencil. His friend bought 2 pens and 3 pencils for 17 zeds. How many zeds will Joe need to buy 1 pen and 2 pencils?

Show your work.

<table>
<thead>
<tr>
<th>Code</th>
<th>Response</th>
<th>Item: M042263</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Correct Response</strong></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10 zeds and equation(s) shown. Equations should involve the use of letter(s) as variable(s), e.g., (2y + 3x = 17).</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>10 zeds and other work shown, e.g., pen = pencil + 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Incorrect Response</strong></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>10 zeds, no work shown</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>Other incorrect (including crossed out, erased, stray marks, illegible, or off task)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Nonresponse</strong></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>Blank</td>
<td></td>
</tr>
</tbody>
</table>

Notes. 1. A correct response is coded with the numbers 10 and 11. The different codes are used to score responses students might give. Incorrect is coded as 70 and 79, and a nonresponse is coded with a 99.
2. Obtained from Mullis & Martin, 2011d.
Figure C.3. TIMSS: Two Point Constructed Response, Eighth Grade.

What is the sum of all the interior angles of pentagon ABCDE? Show your work.

Answer: ______________

Note: Units not required provided correct units implied by the work shown

<table>
<thead>
<tr>
<th>Code</th>
<th>Response</th>
<th>Item: M032692</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Correct Response</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>540 degrees with work shown</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Examples:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 (triangles) × 180° = 540°</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6 (right angles) × 90° = 540°</td>
<td></td>
</tr>
<tr>
<td><strong>Partially Correct Response</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>540 degrees with no work shown</td>
<td></td>
</tr>
<tr>
<td><strong>Incorrect Response</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>Incorrect (including crossed out, erased, stray marks, illegible, or off task)</td>
<td></td>
</tr>
<tr>
<td><strong>Nonresponse</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>Blank</td>
<td></td>
</tr>
</tbody>
</table>

Notes. 1. A correct response is coded with the number 20. The different codes are used to score responses students might give. A partially correct response is coded with a 10. Incorrect is coded as a 79, and a nonresponse is coded with a 99.

Appendix D: TIMSS Test Item Block Distribution

Table D.1. TIMSS: 2007 and 2011 Fourth Grade and Eighth Grade Test Item Block Distribution.

<table>
<thead>
<tr>
<th>Student Achievement Booklet</th>
<th>Mathematics Assessment Blocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Booklet 1</td>
<td>M01</td>
</tr>
<tr>
<td>Booklet 2</td>
<td>M02</td>
</tr>
<tr>
<td>Booklet 3</td>
<td>M03</td>
</tr>
<tr>
<td>Booklet 4</td>
<td>M04</td>
</tr>
<tr>
<td>Booklet 5</td>
<td>M05</td>
</tr>
<tr>
<td>Booklet 6</td>
<td>M06</td>
</tr>
<tr>
<td>Booklet 7</td>
<td>M07</td>
</tr>
<tr>
<td>Booklet 8</td>
<td>M08</td>
</tr>
<tr>
<td>Booklet 9</td>
<td>M09</td>
</tr>
<tr>
<td>Booklet 10</td>
<td>M10</td>
</tr>
<tr>
<td>Booklet 11</td>
<td>M11</td>
</tr>
<tr>
<td>Booklet 12</td>
<td>M12</td>
</tr>
<tr>
<td>Booklet 13</td>
<td>M13</td>
</tr>
<tr>
<td>Booklet 14</td>
<td>M14</td>
</tr>
</tbody>
</table>

Notes.

1. Each student booklet has mathematics and science test blocks. In some test booklets, science blocks are presented to students first. In others, mathematics blocks are presented first. The above table only provides information about which mathematics test blocks student are given according to student booklet number.

2. TIMSS 2007 information adapted from Mullis, Martin, Ruddock, O’Sullivan, Arora, & Erberber, 2005.

Appendix E: Exploratory Analyses

Prior to the factor analysis, an exploratory correlation and regression analysis were conducted using 1st plausible values and self-efficacy (as a continuous variable). The objective of this work was to explore whether or not there was a relationship between mathematics means and self-efficacy prior to additional work using 5 plausible values and prior to grouping of self-efficacy values into high, medium, and low categories. This analysis was conducted using SPSS and 2007 fourth grade and 2011 eighth grade TIMSS data.

The results of the work are displayed in Table E.1, Figure E.1, and Figure E.2. In Table E.1, fourth and eighth grade data show a meaningful correlation of .43 between the 1st plausible values and self-efficacy.
Table E.1. Grade 4 and Grade 8 TIMSS Data: Correlation of 1st Plausible Values and Self-Efficacy

<table>
<thead>
<tr>
<th></th>
<th>Pearson Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIMSS 2007, Grade 4</td>
<td>.425</td>
</tr>
<tr>
<td>TIMSS 2011, Grade 8</td>
<td>.434</td>
</tr>
</tbody>
</table>
A regression analysis provided variances ($R^2$) of .18 or .19\textsuperscript{9}. The linear regression equations show that boys’ means scores increase above girls’ by 42 to 43 points when self-efficacy is factored in. The linear regression equation, line of best fit, and variances are presented in figures E.1 and E.2.

The correlation and regression analysis both provide evidence of a relationship between self-efficacy and 1\textsuperscript{st} plausible values. This warrants a further investigation into the relationship and predictability of self-efficacy to mathematics achievement. To do this, self-efficacy will be explored at high, medium, and low levels to examine how varying levels of self-efficacy affect mathematics achievement. For more information on the methods used to explore this, please refer to Chapter 3.

\textsuperscript{9} Cubic and quadratic regression equations yielded similar variances, .18 for the fourth grade and .19 for the eighth.
Notes.

1. \( R^2 = .181 \)
2. Regression equation: \( 396 + (43 \times \text{self-efficacy}) \)
Notes.

1. \( R^2 = .189 \)
2. Regression equation: \( 385 + (42 \times \text{self-efficacy}) \)