Title
School Variability and Associated Factors in Within-school Gender Differences in Mathematics Performance

Author
Yun, Jung-eun Ellie

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

School Variability and Associated Factors in Within-school Gender Differences in Mathematics Performance

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Education

by

Jung-eun Yun

2019

ABSTRACT OF THE DISSERTATION

School Variability and Associated Factors in

Within-school Gender Differences in Mathematics Performance

by

Jung-eun Yun

Doctor of Philosophy in Education

University of California, Los Angeles, 2019

Professor Noreen M. Webb, Co-Chair

Professor Michael H. Seltzer, Co-Chair

Over the past few decades, education researchers have repeatedly found that male students outperform their female peers on standardized mathematics tests. The attention has turned to factors associated with this phenomenon, with student characteristics and social contexts being a prime focus. Less studied, however, is variability in the magnitude of gender differences in mathematics performance across schools and the impact that school environments might play in this gender difference. The current study seeks to: a) investigate whether differential gender performance exists in a large-scale mathematics assessment and to what extent gender differences vary across schools and b) identify school factors associated with school variability in the size of gender differences in mathematics performance.

The current study uses data from the Trends in International Mathematics and Science Study (TIMSS) collected in 2015 from 8^{th} graders and high performing 12^{th} graders in the US. Hierarchical Linear Modeling (HLM) is the main statistical analysis framework for the study. In particular, this study uses a two-level HLM models with students as the level I model unit and schools or teachers as the level II model unit.

The results indicate that for students in both 8^{th} and 12^{th} grades, males perform better than females. There was also substantial variability between schools in the size of the gender difference in mathematics performance. School-level HLM analyses indicate several predictors are associated with the within-school gender gap, such as the presence of a mathematics resource shortage, communication among teachers, teacher confidence, and teacher support of student participation. For high performing12^{th} graders, whether teachers felt their school was safe and orderly was associated with the within-school gender gap in mathematics performance.

Further analyses using a teacher-level HLM demonstrate that female teachers and teachers under age 30 are more likely to have classes with a wider gender gap in mathematics performance. In addition, and consistent with results from the school-level HLM, teachers who support student participation are more likely to have classes with a smaller gender gap in mathematics performance. Altogether, this study shows the complexity of the within-school gender gap in mathematics performance and suggests the need for future studies.

The dissertation of Jung-eun Yun is approved.

James W. Stigler

Minjeong Jeon

Noreen M. Webb, Committee Co-Chair

Michael H. Seltzer, Committee Co-Chair

University of California, Los Angeles

2019

To my mother

Table of Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my advisor, Professor Mike Seltzer for his guidance, mentorship, patience, and support throughout this incredible journey. He went above and beyond to teach me the joy of learning and conducting research, which I will take with me wherever I go. I would also like to thank Professor Noreen Webb for her wisdom and support in the process of my research. I am grateful to the other members of my committee - Professor Minjeong Jeon and Professor Jim Stigler - for their helpful comments and suggestions. I would also like to thank Professor John Rogers for providing me with the invaluable opportunity to learn from him as a research assistant. I am also indebted to my friends, Kayla and Hye-jin - this work could not have been completed without their unfailing support. . Finally, I would like to express my deepest gratitude to my husband, John, my son, Ian, and my family in Seoul for their enduring love and support throughout. Thanks to you all, it has truly been a wonderful journey.

VITA

Jung-eun (Ellie) Yun

Education

University of California, Los Angeles
Ph. D. Candidate, Education, Social Research Methodology Division, June 2017

University of Washington, Seattle
M.S., Psychology, June 2012

Yonsei University, Seoul, Korea

M.A., Psychology, August 2008

B.A., Psychology & English Language and Literature, February 2003

Publications (selected)

Rogers, J., Franke, M., Yun, J., Ishimoto, M., Diera, C., Geller, R., Berryman, A., & Brenes, T., (2017). Teaching and learning in the age of Trump: Increasing stress and hostility in America's high schools. Los Angeles, CA: UCLA's Institute for Democracy, Education, and Access.

Rogers, J., Mirra, N., Seltzer, M., & Yun, J. (in progress). Unequal Time: Learning Time and Social Class in in U.S. Secondary Schools.

Lee, Y., Yun, J. E., Kim, E., & Song, H. (2015). The development of infants' sensitivity to behavioral intentions when inferring others' social preferences, PLoS ONE, doi:10.1371/journal.pone.0135588.

Rogers, J., Mirra, N., Seltzer, M., & Yun, J. (2014). It's About Time: Learning Time and Educational Opportunity in California High Schools. Los Angeles: UCLA IDEA.

Sommerville, J. A., Schmidt, M., Yun, J., & Burns, M. (2012). Fairness expectations and prosocial behavior in 12-month-old infants, *Infancy*, doi: 10.1111/j.1532-7078.2012.00129.x.

Yun, J., & Song, H. (2007). The development of abilities to infer others' dispositions in preschoolers. *The Korean Journal of Psychology, 26,* 85-102.

Song, H., & Yun, J. (2007). The influence of discourse information on the comprehension of null-subject sentences in Korean adults. *The Korean Journal of Experimental Psychology, 19,* 263-280.

Conference Presentations (selected)

Seltzer, M. & Yun, J. E. (2015). Investigating factors underlying the variation in instructional time loss within and between high schools in California serving communities with high concentrations of poverty: A multilevel analysis. American Educational Research Association Annual Meeting, Chicago, IL.

Mirra, N., Rogers, J., Yun, J. E., & Seltzer, M. (2014). Cracks in the hourglass: Inequalities in learning time between low-and high-poverty high schools. American Educational Research Association Annual Meeting, Philadelphia, PA.

Yun, J., Burns, M., & Sommerville, J. A. (2012). The effect of exposure on infants? Perception of the goal-directed actions of other-race individuals. International Conference on Infant Studies, Minneapolis, MN.

Yun, J., & Sommerville, J. A. (2010). Infants' understanding of others' behaviors: Do infants tie goals to individuals? International Conference on Infant Studies, Baltimore, MD.

Blumenthal, E., Yun, J., & Sommerville, J. A. (2009). Infants' generalization of an actor's goals during the first year of life. Biennial Meeting of Cognitive Development Society, San Antonio, TX.

Yun, J., & Song, H. (2008). Infant's ability to infer agents' dispositions based on underlying intentions. International Conference on Infant Studies, Vancouver, Canada.

Honors and Scholarships

Dissertation Year Fellowship (2018-2019), University of California, Los Angeles

Graduate Summer Research Mentorship Fellowship (2015), University of California, Los Angeles

Graduate Division Fellowship (2013-2017), University of California Los Angeles

Yonsei International Foundation Scholarship (2012), Yonsei International Foundation

*Excellent Thesis Awa*rd (2008), Graduate School of Yonsei University, Korea

Brain Korea 21 Scholarship (2007-2008), Ministry of Education & Human Resources Department, Korea

Scholarship for Student Excellence (2006), Graduate School of Yonsei University, Korea

Scholarship for Academic Excellence (1998-2001), Undergraduate School of Yonsei University, Korea

Chapter 1. Introduction

In an age where technology is flourishing at an exponential pace (Butler, 2016), the influence of STEM fields (science, technology, engineering, and mathematics) is more prevalent than ever. According to the U.S. department of Commerce, STEM occupations are growing at nearly double the rate compared to all other fields[1]. However, according to the World Economic Forum, only 26% of the STEM workforce in developed countries is represented by women[2]. This gender disparity in STEM occupations is disconcerting, as women received less than 20% of the bachelor's degrees in computer science or engineering, even though a majority of undergraduate students are female (NSF, 2016). Not only is a lack of gender diversity in STEM harmful for innovation[3] (Botella, Rueda, López-Iñesta, & Marz, 2019), it may also exacerbate gender income inequality, as STEM jobs are often high-paying jobs. To better understand the gender gap in STEM fields, education researchers and policy makers have focused on the K-12 educational setting, a critical period in the development of students' career abilities and interests. Mathematics in particular has become an important subject of study, as it is both a primary domain in K-12 curriculum and a fundamental skill in STEM fields.

In K-12 education, it is well established that male students perform better than female students on standardized mathematics tests (e.g, Hedges & Nowell, 1995; Hyde, Fennema, & Lamon, 1990; Maccoby & Jacklin, 1974). Although recent studies have found that the difference between genders in mathematics performance is decreasing, moderate gender differences favoring male students are still reported among top performers (e.g., Ellison & Swanson, 2010;

---

[1] https://www.engineeringforkids.com/about/news/2016/february/why-is-stem-education-so-important-/

[2] https://www.weforum.org/agenda/2016/01/could-women-see-gender-parity-in-the-4th-industrial-revolution/

[3] https://www.weforum.org/agenda/2019/04/business-case-for-diversity-in-the-workplace/

Else-Quest, Hyde, & Linn, 2010). The results from a large-scale standardized mathematics assessment, the U.S. National Assessment of Educational Progress (NAEP), administered from 1990 and 2003, also indicate that the gender gap in mathematics in favor of male students does not diminish across years (McGraw, Lubienski, & Strutchens, 2006).

A number of researchers have investigated various factors associated with gender differences in mathematics achievement[4]. One arm of research has found that there are gender differences in students' attitudes or emotions towards mathematics. For example, past studies have observed that female students are more likely to report higher levels of mathematics anxiety and lower mathematics self-efficacy compared to male students (Devine, Fawcett, Szucs, & Dowker, 2012; Goetz, et al., 2013; Hackett, 1985). Other research has found that contextual factors, such as family and school, also influence students' academic performance (e.g., Pianta, et al., 2008; Whitbeck & Gecas, 1988). Social factors, such as gender-stereotyped beliefs imposed by parents and teachers, may also lead female students to avoid advanced mathematics and believe they are inherently "bad" at the subject.

Although numerous studies have investigated factors associated with gender disparity in mathematics achievement, little research has investigated the impact of school environments and the differences between schools in the size or presence of a gender gap. For example, a school with a supportive teacher network may have a smaller or non-existent gender gap in mathematics performance. It is also possible that teachers who encourage female students to participate in class might promote female students' mathematics achievement by helping them to actively recognize their own knowledge and understanding. Thus, the current study aims to investigate school variability in mathematics gender differences. Indeed, if certain schools have males and

[4] In the current study, mathematics performance and mathematics achievement are used interchangeably.

females performing equally or females outperforming males on mathematics assessments, it would be valuable to understand the associated factors to reduce the gender gap in other schools.

In particular, within-school gender differences may directly influence students' perceptions of gender differences, as their reference groups might be other students in the same school. If a female student attends a school with wide gender differences in mathematics performance favoring males, her perception that males are better than females in mathematics may be stronger than the perception of a female student attending a school with smaller gender differences in performance (or a school in which females outperform males). A wider gender gap in a school that favors males, therefore, might reinforce the gender stereotype that "males do better at math," possibly hindering female students from maximizing their mathematics performance. The perceived gender gap might also affect females' mathematics self-concept, leading them to attribute low mathematics achievement to an inherent lack of capability.

To address these considerations, the current study will investigate variability between schools in gender differences in mathematics performance on an international, large-scale standardized assessment. In addition, factors related to the school and classroom environment will be identified in order to provide specific, actionable ways to help female students improve their performance in mathematics. This study will serve as a stepping-stone for future research that specifically seeks to diminish gender differences in mathematics.

This dissertation is organized as follows. Chapter 2 begins with a review of the existing literature. Chapter 3 presents the data and analytical methods used in the current study. Chapter 4 presents the results on whether there are gender differences in mathematics achievement and to what extent the results vary across schools. Based on these results, Chapter 5 investigates school level variables related to variability between genders in mathematics performance. Chapter 6 takes a closer look at teacher-level variables associated with gender gap in mathematics achievement. Finally, Chapter 7 summarizes the study findings, discusses the study limitations, and suggests directions for future study.

Chapter 2. Review of Literature

Gender differences in mathematics performance have been a long-standing and controversial issue in the field of education. Past studies have found that male students outperform female students in mathematics and males grow mathematics ability faster than female students (e.g, Maccoby & Jacklin, 1974). Hyde, Fennema, and Lamon (1990) meta-analyzed 100 empirical studies published between 1967 and 1987 with a total sample of more than 3 million students to examine gender differences in mathematic performance. They found that while male outperformance in mathematics does not emerge until middle school, moderate gender differences favoring males emerge in high school with an average effect size (*d*) of .29[5]. Male outperformance was also observed among college students (*d*=.41) and adults (*d*=.59). The outperformance of male students was further replicated in another meta-analysis study by Hedges and Nowell (1995). Using representative samples of American adolescents across six separate studies, they found small but consistent gender differences in mathematics performance in favor of male students.

Despite evidence from past studies that males outperform females in mathematics, recent studies with current data have revealed a change in pattern. Hyde and colleagues (2008) analyzed state assessments of cognitive performance administered between 2005 and 2007 from ten geographically diverse states in the United States. The researchers found that the effect size of gender differences ranged between -.02 to .06 with a negligible averaged effect size (*d*=.0065), suggesting that there is no apparent gender difference in mathematics. Another meta-analysis with 242 studies published between 1990 and 2007, representing over 1 million students, also

[5] The effect size *d* was calculated by the average scores for males minus the average scores for females, divided by the pooled within-gender standard deviation.

found no gender difference in mathematics performance (Lindberg et al., 2010). Other studies have even reported that high school female students receive higher grades in mathematics than male students in classroom assessments (Kimball, 1989; Pomerantz, Altermatt, & Saxon, 2002).

However, results from large-scale standardized mathematics tests such as NAEP (National Assessment of Educational Progress) indicate that male outperformance continues to persist (McGraw, Lubienski, & Strutchens, 2006). In particular, some research has found a wider gender gap among high performers (e.g., Hyde & Mertz, 2009; Reis & Park, 2001). For instance, Benbow and Stanley (1980) collected extensive data for the Study of Mathematically Precocious Youth (SMPY) to investigate intellectually talented junior high school students. The researchers selected 9,927 seventh and eighth grade students across six separate rounds between 1972 and 1979 from the top 5-percentile in mathematics ability according to a standardized achievement test and examined their SAT mathematics scores. Gender differences in mathematics scores in favor of male students were consistently observed in all six rounds, with an average standard deviation of 0.4. Also, Benbow and Stanley found a preponderance of male students in the upper tail of the distribution of SAT mathematics scores. Similarly, Ellison and Swanson (2010) investigated the gender gap in mathematics among top performing high school students using data from American Mathematics Competitions (AMC). They found that the male-to-female ratio exceeded 10 to 1 among the top 1% students in AMC, further indicating material gender differences in mathematics performance among high-achieving students.

Male student outperformance has also been reported in the area of solving advanced mathematics problems (e.g., Burton & Lewis, 1996; Mills, Ablard, & Stumpf, 1993). Subsequent research has documented that male students outperform female students in complex and applied mathematics problems which require advanced reasoning and deeper problem-solving skills. For instance, Gallagher and colleagues (2000) found that high school male students were more likely than female students to correctly solve mathematics problems requiring unconventional methods that go beyond traditional strategies learned from textbooks. In contrast, female students outperformed male students in conventional mathematics problems that required computational skills learned from textbooks. Liu and Wilson (2009a) also investigated gender differences among high school students in subdomains of mathematics items using PISA 2000 and 2003 data. They found that male students outperform their female peers in complex multiple-choice items and items in shape and space domains, whereas no apparent gender gap was found in items in the quantity domain. Male outperformance in solving mathematics problems that require higher level of mathematical reasoning and advanced problem-solving ability is problematic in that these are highly desired skills in STEM fields.

Factors Related to Gender Differences in Mathematics

Psychological Factors

A growing body of research has emphasized the impact of student affective and attitudinal factors on gender differences in mathematics performance (Casey, Nuttall, & Pezaris, 1997; Fredricks & Eccles, 2002; Frost, Hyde, & Fennema, 1994). In particular, mathematics self-efficacy, mathematics self-concept, and mathematics anxiety have received considerable attention in studies on mathematics performance and associated gender differences.

Mathematics Self-Efficacy

Mathematics self-efficacy is known as a strong predictor of mathematics performance. Specifically, self-efficacy is defined as “people’s judgments of their capabilities to organize and execute courses of action required to attain designated types of performances” (Bandura, 1986, p. 391). According to expectancy-value theory, individuals’ beliefs about their ability to perform tasks or activities and the extent to which they value the activities explain their choice, persistence, and performance (Atkinson, 1957; Eccles et al., 1983; Wigfield, 1994; Wigfield & Eccles, 1992). People with high self-efficacy choose and pursue tasks that they feel competent in and attribute failures to lack of effort rather than lack of competence (e.g., Schunk & Ertmer, 2000). Students with high mathematics self-efficacy are more likely to engage in challenging mathematics tasks and work harder and persist longer in solving those tasks. A number of empirical studies have documented that positive judgment about self-competence in mathematics is strongly correlated with high performance in mathematics across countries (Chiu & Klassen, 2010; Pajares & Kranzler, 1995; Lent, Brown, & Larkin, 1986; Siegel, Galassi, & Ware, 1985; Pajares & Miller 1997; Turner et al., 2004). In addition, a majority of studies on mathematics self-efficacy have found that female students tend to report lower levels of mathematics self-efficacy than male students (Goetz, et al., 2013; Hackett, 1985; Seegers & Boekaerts, 1996; Reis & Park, 2001). For instance, Junge and Dretzke (1995) examined mathematics self-efficacy of gifted high school students and found that male students reported significantly higher self-efficacy than female students in most of the items measuring mathematics self-efficacy.

Mathematics Self-Concept

Another important attitudinal factor that is related to gender differences in mathematics achievement is mathematics self-concept. Self-concept refers to individuals' general knowledge and perceptions of the self that is continually evaluated by personal inferences about the self (Bong & Clark, 1999). Although self-concept and self-efficacy are both constructs about self-evaluation, self-concept is concerned with the global evaluation of individuals' own skills and abilities, whereas self-efficacy is concerned with individuals' convictions or beliefs about what they can accomplish in given situations. A recent piece of research conducted factor analysis of

Programme for International Student Assessment (PISA) 2003 student questionnaires to investigate whether mathematics self-efficacy and mathematics self-concept are distinct constructs. The results suggest that the two constructs are indeed distinct and that the distinction is present across countries (Lee, 2009). In regards to the relationship between mathematics self-concept and mathematics achievement, past studies have found a positive relationship in a variety of settings across schools (Guay, Marsh, Boivin, 2003; Pajares & Miller, 1994). In addition, researchers have found that male students are more likely to report higher levels of mathematics self-concept than female students (Marsh & Yeung, 1998; Skaalvik & Skaalvik, 2004). For instance, Lindberg and colleagues (2013) examined students' mathematics self-concept and mathematics achievement during the first two years of elementary school, and found that female students' mathematics self-concept was significantly lower than male students in the second grade.

Mathematics Anxiety

Mathematics anxiety is known as a significant impediment to mathematics performance. Richardson and Suinn (1972) defined mathematics anxiety as "a feeling of tension and anxiety that interferes with the manipulation of numbers and the solving of mathematical problems in a wide variety of ordinary life and academic situations." Negative affect towards mathematics lowers students' confidence and motivation to study mathematics, discouraging students from taking advanced mathematics courses and ultimately, the pursuit of scientific careers (e.g., Chipman, Krantz, & Silver, 1992; Hembree, 1990). As a possible mechanism, Ashcraft and Kirk (2001) suggested that high levels of anxiety increase arousal states that drain the resource of working memory, which is crucial for processing mathematics problems. Decreased working memory capacity and functioning potentially hinder students' performance on mathematical problem solving. Past research has also found that mathematics anxiety emerges as early as the first grade (Ramirez, et al., 2012) and increases during junior high school (Hembree, 1990). In particular, female students are more likely to report higher levels of mathematics anxiety than male students, an effect that is observed even among high performing female students (e.g., Devine, Fawcett, Szucs, & Dowker, 2012; Rogers, 2003). For example, Wigfield and Meece (1988) assessed mathematics anxiety of 564 students from 6th through 12th grade and found that female students reported stronger mathematics anxiety than male students. Taken together, this body of research demonstrates that female students regularly report more negative affect and attitude towards mathematics than male students do (Hackett, 1985; Hargreaves, Homer, & Swinnerton, 2008; Lent, Brown, & Larkin, 1986; Randhawa, 1994; Randhawa, Beamer, & Lundberg, 1993; Stipek, 2002).

Social Factors

Although these studies provide compelling evidence of a relationship between mathematics performance and students' affective/attitudinal factors towards mathematics, what is less prevalent in existing literature is research on the moderation effect of social contexts on the relationship between student characteristics and gender differential mathematics performance. Social-cognitive theory posits a constant bidirectional interaction between individual behaviors, personal factors, and environmental influences (Bandura, 1986; Zimmerman, 1989). From this perspective, students' behaviors are interdependent on the surrounding social contexts, such as parents, peers, teachers, and schools, and cannot be explained apart from the influence of social environments. In this sense, the underperformance of female students in mathematics compared to males might be explained by social environments to some extent. In this section, I will introduce two hypotheses to explain the importance of social contexts in students' mathematics performance and follow with a review of studies on parent, community, and school environments that potentially influence gender differential performance in mathematics.

Gender stereotype threat hypothesis posits that exposure to the gender stereotype that "males are better at math" might highlight the gender identity of female students, leading to lower mathematics self-efficacy and self-concept and higher mathematics anxiety (Spencer, Steele, & Quinn, 1999; Steele, 1997). Further research suggests that even unconscious information of gender stereotypes impair female students' mathematics performance (Shih et al., 1999) and that children as young as six years old develop the stereotype that males are smarter than females (Bian, Leslie, Cimpian, 2017). Given that students' gender stereotypes are shaped from their social environments, it is crucial to investigate how social contexts reinforce or weaken gender stereotypes.

Another hypothesis that highlights the influence of social context on gender differences in mathematics performance is gender stratification hypothesis. Baker and Jones (1993) proposed that the societal stratification based on gender, which provides fewer opportunities for females, negatively shapes females' attitudes towards mathematics and leads to poorer performance in mathematics. The researchers suggest that females in societies that provide fewer opportunities to females than their male counterparts perceive weaker links between their achievement and future opportunities, impeding mathematics performance. This hypothesis is partly supported by cross-cultural studies, which found a negative correlation between gender equity index and gender differences in mathematics performance. Else-Quest, Hyde, and Linn (2010) found that cross-national variability in gender differences in mathematics is predicted by some gender equity-related variables such as gender equity in school enrollment, share of research jobs, and parliamentary representation. The hypothesis suggests that female students' mathematics performance may be influenced by their perception of gender stratification within their schools. For example, a disproportionate number of male students in advanced mathematics classes may reinforce female students' perceptions that they have fewer opportunities to succeed in mathematics, further impairing mathematical performance.

Parental Influences

Parental influence on children's academic performance has received significant attention from researchers (Eccles et al., 1993; Parsons, Adler, & Kaczala, 1982). Several studies have found that parents are primary role models for children and that parents' academic values are transmitted to their children (e.g., Whitbeck & Gecas, 1988). Eccles and colleagues (1982), for instance, observed that parents' beliefs about the importance of mathematics were positively correlated with their children's values and attitude towards mathematics between grades 5 and 11. Furthermore, when parents hold gender-related stereotype beliefs in mathematics, they may influence their children's mathematics self-concept, exacerbating gender differences in mathematics performance. Related studies have found that parents tend to expect boys to perform better than girls at mathematics and encourage boys more than girls to achieve higher grades in mathematics (Yee & Eccles, 1988). Also, children whose parents hold gender-related stereotype perceptions are more likely to perceive themselves in terms of gender stereotypes (Eccles, Jacobs, & Harold, 1990; Tiedemann, 2000). These empirical findings suggest that parental beliefs and expectation are important social environmental factors that may contribute to gender differential performance in mathematics.

Influence of the Surrounding School Community

The communities surrounding students serve as crucial socialization and learning contexts. Some research suggests that the impact of community factors, such as urbanicity and affluence, on students' academic performance might not be uniform across gender. Bessudnov and Makarov (2013), for instance, investigated community factors related to gender differences in mathematics performance of Russian high school students and found that school location is related to gender differences. In urban schools, male students perform better than female students, whereas schools in rural areas do not show gender differences. Entwisle, Alexander, and Olson (1994) also found the level of affluence in school neighborhoods to be positively linked to mathematics performance of male students, while no link was found for female students. They argue that affluent neighborhoods provide various extracurricular activities that encourage male students to explore new opportunities. These out-of-curriculum-based activities might help male students develop additional problem-solving abilities and strategies, leading them perform better at advanced mathematics problems and potentially widen the gender gap. Legewie and DiPrete (2014) also demonstrated that gender segregation of extracurricular activities might reinforce female students' gender stereotypes, discouraging them from pursuing STEM fields.

Effect of Teachers and Instruction

Existing research on the impact of school climate on student achievement focuses primarily on teacher effect (e.g., Goodenow, 1993; Midgley et al., 1989; Skinner & Belmont, 1993). Teacher support is a topic that has attracted a great deal of research in education. Supportive teachers show a commitment to students' learning and provide a caring and understanding environment. A large body of research has shown that students' perception of teacher support is associated with a wide range of positive outcomes of students. Students who perceive their teachers as supportive are more likely to have high self-efficacy (Fast, et al., 2010; Murdock & Miller, 2003; Patrick, Ryan, & Kaplan, 2007), high motivation and school engagement (Ryan & Patrick, 2001), and high academic performance (Pianta, et al., 2008).

Moreover, the effect of perceived teacher support is moderated by the gender of students. For instance, Goodenow (1993) found that the perceived teacher support is more highly correlated with female students' positive outcomes than those of male students. Also, Reddy, Rhodes, and Mulhall (2003) indicate that perceived teacher support is a stronger predictor of female students' self-esteem and depression than those of male students. Thus, the differential effect of teacher support by student gender suggests that social environmental aspects need to be investigated by gender as well.

Another well-studied teacher effect is how teachers deliver instruction in classrooms, which is directly related to students' academic performance. Teachers who provide cognitively challenging tasks, accompanied with appropriate feedback, activate students' cognitive abilities by encouraging them to think more deeply about finding solutions. Research has found that high school students who perceive their classroom as cognitively challenging showed higher academic self-efficacy (Gentry & Owen, 2004). Also, teachers who set clear goals for mastery and understanding, together with feedback on students' progress, help students improve academic performance (e.g., Fuchs, Fuchs, Hamlett, & Allinder, 1991). In contrast, teachers who emphasize performance goals and encourage competition among students often lead students to use shallow learning strategies (Meece, Blumenfeld, & Hoyle, 1988) and avoid challenging tasks to protect their self-esteem (Dweck & Leggett, 1988). Although effective instruction improves both male and female students' academic performance, gender differential classroom behaviors and attitudes suggest that the extent to which the effective instruction benefits performance might differ by gender. Ma and Willms (2004) found that 8^{th} grade US female students are more likely to have a closer relationship with teachers, be more sensitive to disciplinary school climates, and conform more to school rules than their male counterparts. Moreover, female students are more sensitive to classroom environments than male students and are less disruptive and more engaged in school than male students (Schunk & Lilly, 1984). This female student tendency to conform to rules and engage more deeply in class may allow them to benefit more from effective teacher instruction than male students do. In this sense, it is possible that the effectiveness of a teacher in mastery goal setting, clear feedback, and instruction might affect students' academic performance differently based on gender.

School Variability in Gender Differences in Mathematics

Past research has examined possible school factors associated with gender differences in mathematics, but little research has directly investigated school variability in gender differences in mathematics. Gender differences in mathematics might not be uniform across schools. That is, some schools might have wide gender gaps, whereas other schools do not show any gender differences. Since students' sense of self and the world viewpoint grow within a school environment, within-school gender differences in mathematics performance may directly affect students' perceived gender gap, potentially influencing their mathematics performance. For instance, a female student attending a school with a wider gender gap in mathematics might reinforce gender stereotypes in mathematics due to the perceived wider gender gap. The heightened perception that males are better than females at math may lower the female student's mathematics self-efficacy and self-concept and increase her mathematics anxiety, possibly leading the student to attribute low mathematics scores to a lack of mathematics ability. In contrast, a female student attending a school with no gender gap might not be affected by the gender stereotype, resulting in the attribution of poor mathematics performance to lack of effort. In particular, female students who do not perceive a gender gap in mathematics may maintain a positive outlook on mathematics, enabling them to receive the full benefit of effective instruction and supportive teachers. Thus, focusing on school variability in gender gaps in mathematics performance is an important topic of research and investigating factors that contribute to the school variability is key to understanding and addressing gender differences in mathematics.

One study examined within-school gender differences in academic performance of high school students. Using cross-national PISA 2000 data, Ma (2008) examined the within-school gender gap in 41 countries for three domains: reading, mathematics, and science. He found that school factors associated with gender differences in academic performance are domain-specific and country-specific. Using multilevel modeling, he found that schools in the United States show considerable variation in terms of gender gaps in performance for all domains. The results show that among various school-level factors, teacher shortage is the only significant factor affecting within-school gender gaps in mathematics among schools in the United States. Schools with a high level of teacher shortage showed larger gender differences in mathematics performance favoring male students. However, because of the cross-national nature of the study, the study does not provide sufficient information on country-specific results. The shortage of teacher variable was measured by asking school principals whether the school's capacity to provide instruction is hindered by a lack of qualified mathematics teachers. A lack of qualified teachers is usually associated with school poverty level, but Ma's study did not investigate whether principals' perceptions of teacher shortage is due to school poverty or some other factors. Since the results section did not show how other proxy measures of school poverty level are associated with within-school gender gap in mathematics, it is difficult to conclude whether the shortage of teachers or school poverty levels is the most important factor explaining US school variability in gender differences in mathematics performance. Moreover, since the main domain of PISA 2000 was reading, the dataset is more appropriate for investigating reading gender gap rather than mathematics or science gender gap due to the limited assessment of mathematics and science domains. The present study will use the recent data and further investigate school level factors associated with school variability in gender differential mathematics performance among secondary schools in the United States.

To summarize, there is extensive research that has investigated various students and social factors associated with gender differences in mathematics performance. However, most of the literature has focused on general gender differences in mathematics performance and attended less to school variability in gender differential mathematics achievement. Thus, additional research is needed in these areas to more thoroughly explicate the gender differences in mathematics and find new ways to reduce the gender gap in mathematics. The current study investigates school variability in gender gap in mathematics among US schools and associated school factors using recent data. By elucidating how school level factors affect students' mathematics performance, the present study aims to contribute to the reduction in gender differences in mathematics achievement.

The Current Study

The current study addresses the following questions concerning gender differences in mathematics performance using TIMSS (Trends in International Mathematics and Science Study) 2015 mathematics assessment data on US students.

Question 1. Are there gender differences in mathematics achievement after controlling for students' SES? To what extent do gender differences in mathematics achievement vary across schools?

Using TIMSS' recent large-scale standardized international assessment data, I examined whether there are gender differences in mathematics achievement and investigated school variability in gender differences. By using hierarchical linear modeling, I investigated whether there is a significant amount of variation in within-school gender gap across schools. Specifically, I conducted an assessment for each of the two different datasets: one assessed regular 8^{th} grade students' mathematics achievement and the other assessed high achieving 12^{th} grade students' mathematics performance.

Question 2. What school-level factors are associated with school variability in gender differences in mathematics achievement?

Research question 1 found that some schools have wider gender differences than other schools. Given the variability of gender gap between schools, I investigated school level factors that account for school variability in gender differences in mathematics achievement. I examined question 2 using two different models: school-level HLM and teacher-level HLM. The first model concerns the general relationship between within-school gender gap and school-level factors. Since some of important factors were constructed based on teacher questionnaires, I built the second model to closely investigate the relationship between teacher-level variables and the gender gap in mathematics performance.

The findings from the above research questions will provide further insight into explaining gender differences in mathematics performance.

Chapter 3. Methods

The TIMSS Data

To address the research questions, this study entails a secondary-analysis of the Trends in International Mathematics and Science Study (TIMSS) mathematics assessment data. In particular, I analyzed TIMSS 2015 mathematics assessment data for eighth grade students (TIMSS 2015 Regular) and TIMSS 2015 Advanced mathematics assessment data for high performing twelfth grade students (TIMSS 2015 Advanced) in the United States.

There are several advantages to using TIMSS 2015 Regular and TIMSS 2015 Advanced data for investigating gender differences in mathematics achievement and associated school variability. First, it is an international, large-scale standardized test with a representative sample of the population. Indeed, its rigorous sampling design reflects the hierarchical structure of the education system, enabling researchers to examine not only characteristics of the students, but characteristics of the schools they attend. Second, the TIMSS 2015 Regular and TIMSS 2015 Advanced assessments provide data across two different populations. Eighth grade students who took the TIMSS 2015 Regular assessment were sampled to represent the overall population of 8th grade students in the US. On the other hand, 12th grade students who took the TIMSS 2015 Advanced assessment were sampled to represent high performing 12th grade US students in mathematics. Though past research has found wider gender differences among high school students and among high performers in mathematics (e.g., Hyde & Mertz, 2009; Reis & Park, 2001), the current study will be able to better investigate the potential reasons behind the gender gap by probing some potential factors (e.g., school characteristics, teacher instruction, etc.). The third advantage to using TIMSS data for the present study is that the TIMSS 2015 Regular and

Advanced assessments contain context questions for students, teachers, and school principals. Context questions provide information on students' psychological and sociological environments which may help to explain gender differences in mathematics performance. Lastly, the data is relatively recent, generated in 2015, so it can speak to the current state of mathematics gender differences and variability across schools.

TIMSS 2015 Regular

TIMSS is an internationally standardized assessment of mathematics and science achievement administered at fourth and eighth grades. Since 1995, TIMSS has been administered every four years by the International Association for the Evaluation of Educational Achievement (IEA). The primary goal of TIMSS is to assess student performance in mathematics and science and to provide important background information, such as instructional practices, curricula, home and school environment, with the aim of improving teaching and learning in mathematics and science. The present study used mathematics assessment data from eighth graders in the USA in 2015 to examine the current state of gender differences in mathematics.

TIMSS 2015 Regular has two dimensions: a content domain and a cognitive domain. In 2015, the TIMSS 2015 mathematics assessment included four areas in the content domain: *number*, *algebra*, *geometry*, and *data and chance*. The number area includes comprehension and skills related to 1) whole numbers, 2) fractions, decimals, and integers, and 3) ratio, proportion, and percent. Algebra consists of 1) expression and operations, 2) equations and inequalities, and 3) relationships and functions. The geometry area assesses students' understanding of geometric relationships, including 1) geometric shapes, 2) geometric measurements, and 3) location and movement. Finally, the data and chance area contains 1) characteristics of data sets, 2) data interpretation, and 3) chance.

The cognitive domain of TIMSS assesses mathematical thinking processes including *knowing*, *applying*, and *reasoning*. *Knowing* focuses on mathematical knowledge about facts, concepts, and procedures, whereas *applying* involves applying knowledge to solve mathematical problems. *Reasoning* assesses logical and inductive thinking based on prior knowledge to draw solutions for novel or unfamiliar problem situations.

TIMSS 2015 Regular contains two types of items: multiple choice and constructed-response. Constructed-response items ask students to construct a written response. Multiple choice items are worth approximately half of all points (each item is worth one or two points, depending on the degree of complexity). To score mathematics assessment items, the TIMSS administration used item response theory (IRT). For multiple choice items, a three-parameter IRT model was used, while constructed-response items used either a two-parameter model or a partial credit model, depending on the number of answering options. A more detailed description of the IRT model will be briefly explained in the "overview of analysis methods" section.

To ensure a representative sample of the population, TIMSS adopted a two-stage cluster sampling procedure. In the first sampling stage, schools were systematically selected with probabilities proportional to the number of students enrolled. The second sampling stage involved randomly sampling one or more intact classes of eighth grade students within each selected school. In the United States, two classes were randomly selected from most of the selected schools. All students in each selected class participated in the mathematics assessment.

All in all, the mathematics and science assessment took 90 minutes and was completed with paper-and-pencil. After the test, students were asked to complete a 30 minute-long questionnaire covering contextual information on family and school. Questionnaires were also administered to school principals and mathematics teachers of those students in order to gather further contextual information. In the 2015 cycle, students from 57 countries participated in the TIMSS assessment. In the United States, 10,221 eighth grade students from 246 schools participated (50.1% females).

TIMSS 2015 adopted a rotated block design in order to minimize the burden of students and to broadly cover subject contents to estimate proficiency in populations. Thus, students were randomly assigned to one of fourteen test booklets consisting of clusters of mathematics and science assessment items. Because all students were not given the same cognitive assessment items, a student's raw score in mathematics and science does not provide an accurate estimate of an individual student's proficiency in mathematics and science. As such, the TIMSS administration used the multiple imputation method to generate scores for students' proficiency. The multiple imputation process entails estimating students' proficiency, given students' responses to items and their background data. However, imputing and assigning a single value for each individual does not account for imputation error and the uncertainty of estimation. Thus, TIMSS generated five plausible values for each student to estimate their proficiency in order to incorporate uncertainty into individual students' scores. The five plausible values for each student were selected from the conditional distribution that could be reasonably assigned to each individual. Thus, the five plausible values represent the range of scores a student might achieve if he or she had completed the entire assessment (TIMSS & PIRLS International Study Center, 2016).

Multiple imputation was also used for estimating proficiency in each sub domain. Thus, nine sets of five plausible values (overall mathematics score, *number*, *algebra*, *geometry, data and chance*, *knowing, applying,* and *reasoning*) were assigned to each student.

TIMSS 2015 Advanced

TIMSS 2015 Advanced was administered for students enrolled in advanced mathematics and/or physics classes in the final year of secondary school (12^{th} grade). TIMSS advanced was administered in 1995, 2008, and 2015 (the US participated in the 1995 and 2015 administrations). An “advanced mathematics class” is defined as a calculus course in the US data.

The TIMSS 2015 Advanced includes three areas in the content domain: *algebra, calculus*, and *geometry*. Consistent with the TIMSS 2015 Regular assessment, the cognitive domain of the TIMSS 2015 Advanced consists of *knowing*, *applying*, and *reasoning*.

To ensure a representative sample of the population, during the first stage, schools were selected systematically with probabilities proportional to the number of 12^{th} grade students who have taken courses in advanced mathematics. The second stage involved randomly sampling 12^{th} grade students who take advanced mathematics classes (i.e., calculus) within each selected school. In the 2015 cycle, students from nine countries participated in the TIMSS Advanced assessment. In the US, 2,954 twelfth grade students participated, from 241 schools (51% females).

Similar to TIMSS 2015 Regular, six booklets of mathematics assessments were randomly assigned to students using a rotated block design, in order to minimize students’ burden. Five plausible values were provided for each individual, based on the IRT model, to estimate their proficiency in mathematics.

Overview of Analysis Methods

Hierarchical Linear Regression

Hierarchical linear modeling (HLM) is the main statistical analysis framework for the study. HLM was specifically developed for hierarchically structured data, like the data analyzed in this paper. In HLM, students are nested within higher-level, social environment units such as classrooms and schools. These higher-level units of social environments can potentially impact students' outcomes, so students in the same classroom or school may be more similar than students from different classrooms or schools.

There are a couple of problems with using a single-level ordinary least squares (OLS) regression when analyzing hierarchically structured data. For one, similarity among students nested within the same group violates the single-level, OLS regression assumption - independence of observation. Thus, using a single-level OLS regression with hierarchically structured data results in biased estimates of parameters of interest. Another problem with single-level OLS regression for the present data is that group effects on individual outcomes, which can differ across groups, are overlooked.

HLM is a powerful and flexible method for analyzing nested data. HLM enables researchers to investigate both individual and group level effects by building up separate models for each level and estimating the multiple levels simultaneously. With multilevel models, HLM takes into account the heterogeneity in relationships of interest across groups by incorporating variables that are conceptually defined at different levels. It also allows researchers to investigate the effects of group-level variables, individual level variables, and even interaction effects between individual and group variables. Thus, HLM is an appropriate analysis tool for investigating school variability in the gender gap of mathematics performance and factors related to this variability.

The HLM analyses in the current study include two-level HLM models with students as the level I model unit, and schools (Chapter 5) or teachers (Chapter 6) as the level II model unit. This two-level HLM will examine the extent to which math performance differs between genders as a function of school, and whether school variation, if any, is systematically related to the school (or teacher) level variables. In addition, HLM can investigate cross-level interactions between student characteristics and school (or teacher) level variables.

The first HLM analysis (Chapter 4) will investigate the extent to which schools vary in terms of the gender difference in each school in mathematics performance and subsequent questions/chapters will explore school/teacher factors that are related to gender differences. In particular, a second HLM analysis (Chapter 5) will investigate school factors using a school-level HLM, and a third HLM analysis (Chapter 6) will investigate school factors using a teacher-level HLM. For all of the HLM analyses, individual students' mathematics achievement scores are modeled as a linear relationship of students' gender and students' SES at the student level (level 1). In the first HLM analysis, no predictors are specified at the school level (level 2). In the second HLM analysis, school-level variables are included in the school level (level 2) to investigate school-level variables that are associated with the within-school gender gap in mathematics achievement. In the third HLM analysis, teacher-level variables are specified in the teacher level (level 2) to more closely investigate teacher-level variables that are associated with the within-school gender gap in mathematics achievement.

Software developed to analyze hierarchically structured data, HLM 7.0, was used for all HLM analyses (Raudenbush, Bryk, Cheong, & Congdon, 2011), as it is capable of dealing with complex sampling design weights at different levels of analysis and incorporating five plausible values as outcomes in the analyses.

Weight

TIMSS 2015 used a two-stage cluster sampling procedure, yielding the sampled students and schools that do not have equal probability of selection. Because of the unequal probability of selection, biased parameter estimates will be produced when raw data is used for statistical analyses. To avoid this problem, large-scale assessments using a complex sampling procedure provide several weight variables, which are the inverse of the probability of selection, attached to each student and each school. When appropriate weight variables are applied in the analyses, the results return unbiased population estimates to reflect the characteristics of the population (Martin, 2015).

TIMSS 2015 provides several sampling weight variables. Two kinds of weight variables were used in the current study: student weight variables and school weight variables. They contain the probability of selection for each student and school and adjust for non-response (Martin, 2015). The non-response adjustment is included to compensate for sampled students and/or schools who did not participate and were not replaced.

TIMSS provides three different types of student weight variables: the total student weight (TOWGT), the senate weight (SENWGT), and the house weight (HOUWGT). The total student weight variable (TOWGT) in TIMSS has three components that capture the inverse probabilities of selection for (a) the school, (b) classrooms within the school, and (c) individual students within the classrooms (all adjusted for non-participating schools, classes, and students, respectively). It is calculated such that the sum is the student population size in that country. The other two student weight variables were re-scaled from TOWGT. The senate weight (SENWGT) is calculated such that the sum is the sample size of 500 in each country (ignoring actual population size of each country), which is appropriate when researchers conduct cross-national analyses. House weight (HOUWGT) is calculated such that the sum corresponds to the actual sample size, not population size, in the country.

In the multilevel analysis, student weights are applied in the level I model and school weights are applied in the level II model, in order to produce unbiased parameter estimates. However, it is important to note that the TIMSS student weight variable (TOWGT) is calculated as the joint probability of selecting the student, classroom, and school, so using the total student weight (TOWGT) in the multilevel analysis will result in using school weight in both the level 1 and 2 models (Laukaityte & Wiberg, 2018). To avoid this, I created a modified student weight variable (STUDWGT) that contains the probability of just students and classrooms. In the school-level HLM analyses (Chapter 5), modified student weight (STUDWGT) was used for the student level 1 model and school weight (SCHWGT) was used for the school-level 2 model. Both weight variables were normalized so that the sums of each weight were equal to the student and school sample size in the data, respectively.

Centering

In quantitative models, the locations of predictors are related to the interpretation of the intercept and slope parameters. When variables in the models do not have a meaningful zero point, changing the locations of predictors facilitates the interpretation of the intercept and slope. In particular, the choice of centering in hierarchical linear modeling has a material impact on parameter estimates and the interpretations.

There are two main methods for centering in hierarchical linear modeling: *grand-mean centering* and *group-mean centering*. Grand-mean centering expresses predictors as deviations from the grand mean, which is a constant value, whereas group-mean centering expresses the level 1 predictors as deviations from each group mean, which is a group-specific value. In particular, group-mean centering was used in the current study for the student level 1 model and grand-mean centering was used for the school level 2 model. By choosing group-mean centering in the level 1 model and grand-mean centering in the level 2 model, the between-group variation is removed from level 1 and slope heterogeneity across groups can be investigated.

Since a constant overall grand-mean is subtracted from each value of the predictors, grand-mean centering is a simple linear transformation of the uncentered model with the natural metric of predictors.

In contrast, in group-mean centering, the values of predictors entered in the models change depending on the mean of the group/cluster that the individual belongs to. For example, imagine student A in classroom A who has 8 points in mathematics confidence, while student B in classroom B has 9 points. The mean of all the students in classroom A is 6 and the mean of all the students in classroom B is 10. When group-mean centering is used in the HLM, the value of student A becomes 2 (i.e., 8-6) and the value of student B becomes -1 (i.e., 9-10). So, even though student B has a higher raw value, when group-mean centering is applied, student A has a higher value because the mean of classroom B is higher than the mean of classroom A. Thus, the relative standing of the individual within the classroom is the focus in group-mean centering.

Since the two centering methods differ significantly and directly impact the interpretation of parameter estimates, it is important to choose the appropriate centering method for each research question. When investigating cluster effects (level 2 units) while adjusting for level 1 covariates, grand-mean centering is the appropriate choice. In contrast, using group-mean centering in the level 1 model is beneficial in two respects. First, group-mean centering removes the correlation between level 1 and level 2 predictors, providing unbiased estimates of within-school relationships (e.g., gender difference). Second, it is advantageous to obtain unbiased estimates of the extent to which within-group coefficients / slopes vary across schools. Thus, group-mean centering is appropriate when investigating slope heterogeneity for group means that vary significantly across groups.

As this study is primarily investigating the heterogeneity of the gender gap in mathematics performance across schools, group-mean centering was used in the level 1 model.

Item Response Theory – Scoring of Variables

There are two types of variables used in the current study: 1) variables provided by the TIMSS administration and 2) variables that were created by the study author from the TIMSS questionnaire. Both types of variables were scored using an item response theory (IRT) model. IRT is a mathematical, model-based measurement framework in which an individual's underlying construct is estimated based on item characteristics and the individual's responses to those items. For scoring items from a questionnaire or survey, researchers commonly conduct a factor analysis and then create a composite score that weights items according to the factor loadings (i.e., factor scoring). Though factor scoring is a more common method, IRT scoring was selected as more appropriate for the current study. The difference between factor scoring and IRT scoring is that factor scoring treats data as continuous, whereas IRT scoring treats data as categorical (Wirth & Edwards, 2007). As most items in the TIMSS questionnaire have 4-answer categories (e.g., very often, often, sometimes, and never or almost never), I used the IRT scoring method. It should be mentioned that though factor scoring and IRT scoring provide similar results, the IRT method was chosen in order to be consistent with the TIMSS's own scoring method.

In the current study, variables provided by TIMSS administration include 9 sets (overall mathematics score, *number, algebra, geometry, data and chance, knowing, applying,* and *reasoning*) of five plausible values and two additional variables: *instruction affected by mathematics resources shortage* (TIMSS 2015 Regular) and *safe and orderly school* (TIMSS 2015 Advanced). These variables were scored using a two-parameter, three-parameter, or partial credit model, depending on the number of possible answer choices (as explained below).

Three other variables were created by the study author from the teacher questionnaire and scored using the IRT method (graded response model): *teacher confidence in teaching mathematics*, *teacher communication*, and *teacher support of student participation*[6].

Table 3.1. Variables used in the current study and scored by IRT method

Variable	Type of variable	IRT scoring method	Data source
Plausible values for mathematics proficiency	TIMSS provided variable	2-parameter Model 3-parameter Model Partial Credit Model	TIMSS 2015 Regular TIMSS 2015 Advanced
Instruction affected by mathematics resources shortage	TIMSS provided variable	Partial Credit Model	TIMSS 2015 Regular
Safe and orderly school	TIMSS provided variable	Partial Credit Model	TIMSS 2015 Advanced
Teacher confidence in teaching mathematics	Created for the current study	Graded Response Model	TIMSS 2015 Regular
Communication among teachers	Created for the current study	Graded Response Model	TIMSS 2015 Regular TIMSS 2015 Advanced
Teacher support of student participation	Created for the current study	Graded Response Model	TIMSS 2015 Regular TIMSS 2015 Advanced

[6] Tables showing variables in each model is presented in Chapter 4 through Chapter 6.

TIMSS 2015 used the two-parameter model (Bock & Aitken, 1981) to score dichotomous constructed-response items. The two-parameter model predicts the probability that a person will respond correctly to a particular item, given the individual's location on the underlying latent construct, as follows:

$$P(x_{ni} = 1|\theta_n, \alpha_i, \delta_i) = \frac{e^{\alpha_i(\theta_n - \delta_i)}}{1 + e^{\alpha_i(\theta_n - \delta_i)}}$$

where α_i denotes the item discrimination of item i, θ_n indicates student *n*'s location on the underlying latent construct, and δ_i represents item difficulty of item *i*. As implied by the subscript i, items are not equally related to the latent construct and the difficulty of the items vary. The probability of endorsing a particular item depends on the item discrimination (α_i) and the difference between students' location θ_n on the underlying latent construct and the item difficulty ($\theta_n - \delta_i$).

The three-parameter model (Birnbaum, 1968) was used for dichotomously scoring multiple-choice items (correct/incorrect). When multiple choice options are given, even students who are very low on the latent trait have a nonzero probability of correctly responding to items based on random guessing. Thus, the three-parameter model adds a lower-asymptote parameter called a guessing parameter to the two-parameter model so the probability does not equal zero.

The three-parameter model is defined below:

$$P(x_{ni} = 1|\theta_n, \alpha_i, \delta_i, \gamma_i) = \gamma_i + (1 - \gamma_i)\frac{e^{\alpha_i(\theta_n - \delta_i)}}{1 + e^{\alpha_i(\theta_n - \delta_i)}}$$

A lower-asymptote parameter γ_i indicates the probability of correctly responding to item *i* based on random guessing. Scoring of TIMSS 2015 multiple-choice assessment items uses this three-parameter model, incorporating the chance of responding correctly based on random guessing.

In TIMSS data, polytomous constructed-response items and latent constructs from the background questionnaires are scored based on the partial credit model (Masters, 1982). Unlike the two-parameter and three-parameter model with binary data, the partial credit model is one of the IRT models used when data is polytomous. It represents the relation between individual differences on an underlying construct and the probability of responding to an item in a specific category. In particular, the partial credit model posits that the examinee responds sequentially to a number of category options. The term "step" refers to choosing between category options. For instance, if an item has m possible response categories, the examinee must complete m-1 sequential steps in order to respond in the highest category.

The partial credit model (PCM) is shown below:

$$P_{x_i}(\theta_n) = \frac{e^{\sum_{j=0}^{x_i}\left(\theta_n - \delta_i + \tau_{ij}\right)}}{\sum_{h=0}^{m_i} e^{\sum_{j=0}^{x_i}\left(\theta_n - \delta_i + \tau_{ij}\right)}} \quad x_i = 0,1, \ldots., m_i$$

where $P_{x_i}(\theta_n)$ denotes the probability that person n with location θ_n on the underlying latent construct chooses response category x_i on item i out of the m_i possible response categories for the item. The item parameter δ_i denotes the location of the item on the underlying latent construct and τ_{ij} shows the location of step for the response categories (Martin, Mullis, Hooper, Yin, Foy, & Palazzo, 2016).

When responses to the background questionnaire from students, teachers, and principals are scored, those underlying latent variables are not observable and do not have a generic metric. Thus, the constructed latent variable scores are scaled so that the mean is 10 across all countries participating in the TIMSS assessment. Also, the unit of scale is set to 2 points (i.e., the standard deviation across all countries participating in TIMSS).

In addition to the TIMSS derived variables, I created additional variables, such as types of teacher practices and interactions among teachers, based on the responses from students, teachers, and school principals, using a graded response model (GRM; Samejima, 1969; 1996). The graded response model is another type of polytomous IRT model to estimate a person's location on an underlying construct.

A graded response model (GRM) is a generalization of the 2PL model and is based on modeling the process of responding above a between-category boundary (e.g., x = 0 vs. 1, 2, 3; x = 0, 1 vs. 2, 3; and, x = 0, 1, 2 vs. 3). A GRM is an indirect IRT model in that it needs two steps to estimate the probability of each category response. I used a GRM because it is appropriate when item responses are ordered, categorical responses and allows the items to vary in their slope parameters, which represents the degree to which an item discriminates between people with low and high traits. Since most of the item response options for the background questionnaires are ordered by: strongly agree, agree, disagree, and strongly disagree, a GRM is an appropriate method to create TIMSS variables.

The graded response model used in this study is shown below:

$$P^*_{ix}(\theta) = \frac{e^{\alpha_i\left(\theta-\delta_{ij}\right)}}{1+e^{\alpha_i\left(\theta-\delta_{ij}\right)}} \qquad x = j = 0,1,\ldots.,m_i \qquad \text{(a)}$$

$$p_{ix}(\theta) = p^*_{i(j=x)}(\theta) - p^*_{i(j=x+1)}(\theta) \qquad \text{(b)}$$

Separate δ_{ij} parameters are estimated for each step of the item response category and one α_j parameter is used for all steps for each item in the (a) step. $P_{ix}(\theta)$ denotes the probability that person with location θ on the underlying latent construct chooses response category x and above to item i. The item parameter δ_i indicates the between-category threshold and denotes the underlying trait level necessary to respond above threshold j with 0.5 probability. After estimating $P_{ix}^*(\theta)$, actual category response probabilities are calculated in the (b) step by subtraction.

In order to create background variables from the questionnaires, I first conducted exploratory IRT analyses to investigate whether items are clustered as underlying constructs. The cluster of items was assumed to be unidimensional with only one general factor. An exploratory IRT analysis with one factor was conducted to examine whether a one-factor model offers a reasonable explanation of the data. If the factor loadings of those items were reasonably high (more than 0.6), IRT scoring with a graded response model was conducted to obtain a score for each individual for the specific latent constructs. Exploratory IRT analyses were conducted using R's (version 3.5.3) mirt package (Chalmers, 2012) and IRT scoring with graded response models was conducted using SAS (version 9.4).

Chapter 4. Research Question 1

Research Question 1: Are there overall gender differences in mathematics achievement after controlling for students' SES? To what extent do gender differences in mathematics achievement vary across schools?

Data

TIMSS 2015 datasets include observations that are missing in the student, teacher, and school principal questionnaires. In order to estimate parameters of interest, the HLM 7.0 program requires that all level 2 (school-level) variables are valid without any missing values. If there are missing values in one or more level 2 variables, HLM excludes those level 2 units from the analyses. This missing value treatment method is called the listwise deletion, which eliminates cases with any missing values in one or more variables from the analyses.

Since the listwise deletion is used to handle missing data for subsequent analyses, the total sample size depends on the number of missing values of level 1 and 2 variables included in the model. To compare the results from research questions 1 and 2, the same sample was used for both analyses.

For the TIMSS 2015 Regular data, among 246 schools, two schools have only one gender, one of which has all males and the other all females. Since the current study is investigating gender gaps in mathematics performance, I excluded these two schools from the analyses. In addition, 43 schools have missing values in the school-level variables included in the analyses and were omitted from the analyses.

For the TIMSS 2015 Advanced data, among 241 schools, two schools have only female students and three schools have all male students. Those schools were excluded from the analysis. An additional 61 schools were excluded due to missing values on the school-level variables[7].

The total number of schools and students used in addressing questions 1 and 2 are presented below.

Table 4.1. TIMSS 2015 Regular and Advanced Data Sample Sizes

	TIMSS 2015 Regular	TIMSS 2015 Advanced
Number of schools	201	175
Number of students	8,352	2,274
Number of teachers	340	280
Number of girls	4,234	1,166
Number of boys	4,118	1,108

Before conducting analyses, I explored the distribution of both the student weight and school weight variables to see whether there are extreme cases which might cause biased estimates.

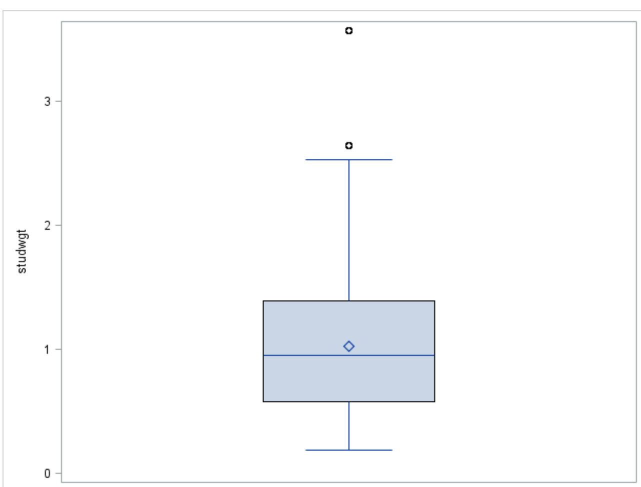


Figure 4.1. Student Weight Distribution for TIMSS 2015 Regular Data

[7] The percentage of schools with missing data on each of the level 2 variables is located in Appendix B (Table B-1 for TIMSS 2015 Regular) and Appendix C (Table C-1 for TIMSS 2015 Advanced)

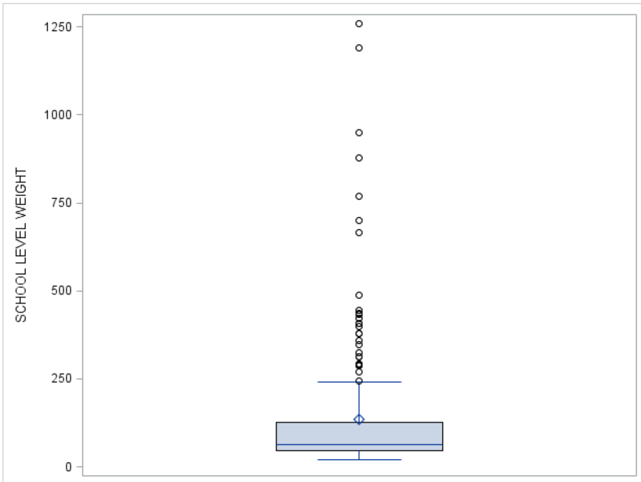


Figure 4.2. School Weight Distribution for TIMSS 2015 Regular Data

Figures 4.1 and 4.2 are the box plots of each variable in the TIMSS 2015 Regular data. As the box plots show, the school weight variables have several extreme values and the student weight has one possible outlier. Because extremely high weight values may potentially impact estimates of parameters of interest, a weight trimming method was used before the subsequent HLM analyses to avoid biased parameter estimates. Weight trimming reduces high weight values to a fixed maximum value (Potter, 1988; 1990). To define a maximum cut-off value, the following formula was used:

Cut-off values = median weight values + 5**Inter Quartile Range (IQR)*

Accordingly, the maximum cut-off value for the student weights is 5.24. The fixed maximum value for school weights is 567.96. Because all student weight values are below the maximum cut-off value, the raw student weight values were used in analyses. For school weight, there were 10 schools with higher than the maximum cut-off value so those schools' weights were trimmed to the maximum value (504.06).

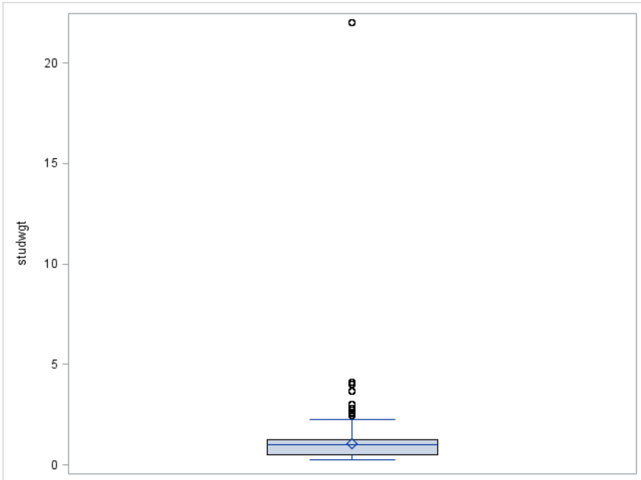


Figure 4.3. Student Weight Distribution for TIMSS 2015 Advanced Data

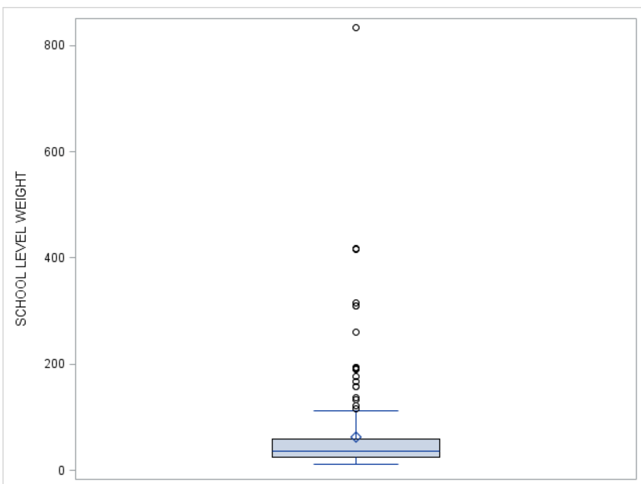


Figure 4.4. School Weight Distribution for TIMSS 2015 Advanced Data

Figures 4.3 and 4.4 are box plots of the student and school weight variables in the TIMSS 2015 Advanced data. The same procedure for the weight trimming was executed for the TIMSS 2015 Advanced data. Accordingly, there were five students whose weight values were greater than the cut-off value (4.76) and eleven schools whose weights were greater than the cut-off value (204.52). Thus, those students and schools weights were replaced by the maximum cut-off weight values (student weight max = 4.10, school weight max = 204.49).

Adjusted weight values using the weight trimming method was used for all subsequent HLM analyses to avoid biased parameter estimates.

Variables of Interest

- *Student gender*: student gender is the main variable. In the models, male was coded as 0 and female was coded as 1.
- *Student Socio-economic status (SES)*: The home educational resources scale, which is provided by the TIMSS administration, was used as a proxy of each individual student's SES. It is a derived variable based on three components: 1) number of books in the home, 2) number of home study supports (i.e., whether students have an internet connection and/or their own room), and 3) the highest level of education of either parent. This TIMSS derived variable is standardized with a mean of 10 and standard deviation of 2 across all participating countries. Higher values represent more home educational resources.
- *Outcome variables*: The main outcome variable is students' overall mathematics performance, as indicated by a set of five plausible values.

To investigate research question 1, five plausible values for each domain (cognitive or content) were used in order to investigate the overall gender gap for each domain and how much mathematics performance varies across domains within schools. Below is the list of outcome variables (each with 5 plausible values) used in research question 1:

- overall mathematics performance

- *algebra* performance (content domain)
- *calculus* performance (content domain)
- *geometry* performance (content domain)
- *data and chance* performance (content domain)
- *knowing* performance (cognitive domain)
- *applying* performance (cognitive domain)
- *reasoning* performance (cognitive domain)

Model Specification

<u>*Null Model*</u>

As a preliminary analysis, a null model was built, in which no predictors were specified at either the student-level or the school-level, to examine the extent to which observations within schools are correlated. The null model is also called a One-Way ANOVA with random effects (Raudenbush & Byrk, 2002). Estimating a null model is a fundamental first step to addressing how much variation in the outcome variable (mathematics achievement) is within or between schools. In other words, the null model enables one to investigate how much similarity there is between students in the same group. The null model is specified below:

Level I: Student level model

$$Y_{ij} = \beta_{0j} + \varepsilon_{ij}, \qquad \varepsilon_{ij}\ N(0, \sigma^2)$$

Level II: School level model

$$\beta_{0j} = \gamma_{00} + u_{0j}, \qquad u_{0j}\ N(0, \tau_{00})$$

In the model, Y_{ij} refers to the mathematics performance score for student *i* in school *j*. Five plausible values for students' mathematics performance were used as outcome variables for the null model. β_{0j} indicates the mean of mathematics performance in school j. ε_{ij} is the unique residual associated with student *i* in school *j* and it is assumed to be normally and independently distributed with a mean of zero and variance σ^2.

In the level II model, γ_{00} refers to the grand mean of mathematics performance across all schools and u_{0j} indicates the deviation of the mathematics score for school *j* from an expected grand mean. τ_{00} indicates the variation in the school mean mathematics score around the grand mean.

The null model is informative in that it partitions the variance in the outcome variable into within and between school variances. The intra-class correlation (ICC), which represents how similar observations are within the same higher level unit compared to how similar mean observations are between different higher level units, can be calculated from the null model.

$$ICC = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

ICC is interpreted as the proportion of variance in the outcome variable that is accounted for by group membership. It also indicates the correlation among observations within the same group. Higher ICC values indicate that individuals in the same group tend to be more similar than groups are from each other.

Model for Question l

An HLM with two levels was built to answer research question 1. In the analysis, SES for each individual student is treated as a covariate since it may potentially influence students' mathematics performance, independent of students' gender. Omitting this covariate in the analysis could result in biased estimates of gender differences in mathematics performance.

At the student level (level I), individual students' mathematics achievement scores are modeled as a linear relationship of students' gender and students' SES.

Level I: Student level model

$$Y_{ij} = \beta_{0j} + \beta_{1j}Gender_{ij} + \beta_{2j}SES_{ij} + \varepsilon_{ij}, \qquad \varepsilon_{ij}\ N(0, \sigma^2) \qquad (1)$$

The level I variables ($SES_{ij}$and $Gender_{ij}$) are expressed as deviations from the school means using group-mean centering in the level I model, since group-mean centering is appropriate for detecting the variability of a within-school gender gap (Raudenbush & Bryk, 2002). The choice of centering affects the interpretation of intercept, but does not affect that the interpretation of the slope (in this context, the gender contrast and SES slope in the model). If grand-mean centering is applied, parameter β_{0j} indicates the adjusted mean mathematics achievement scores for school j, after controlling for gender and SES. By group mean-centering of both gender and SES variables, parameter β_{0j} indicates the average mathematics achievement scores for school j. In the equation above, Y_{ij} indicates the mathematics performance score for student i in school j. $Gender_{ij}$ represents an indicator variable for student i in school j. It is a binary variable with a value of 0 (male) or 1 (female), entered in the equation using group-mean centering. As described above, SES_{ij} indicates students' SES, which is provided by TIMSS to represent students' economic, social, and cultural status based on the student's home educational resources.

The estimated parameter β_{0j} refers to the average mathematics achievement scores for school j. β_{1j}, which is the level 1 parameter of main interest, indicates the expected differences in mathematics achievement scores between male and female students for school j, holding student's SES (SES_{ij}) constant. β_{pj} is the slope of predictor SES_{ij}, which indicates the relationship between students' SES and mathematics achievement scores within school j. ε_{ij} is the level 1 residual term that is assumed to be normally and independently distributed with mean zero and variance σ^2.

Next, a Level II model was posed to represent the between-school relationship.

Level II: School level model

$$\beta_{0j} = \gamma_{00} + u_{0j}, \qquad (2)$$

$$\beta_{1j} = \gamma_{10} + u_{1j}, \qquad (3)$$

$$\beta_{2j} = \gamma_{20} + u_{2j}, \qquad (4)$$

$$\mathrm{Var}\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{bmatrix} = \begin{bmatrix} \tau_{00} & \tau_{01} & \tau_{02} \\ \tau_{10} & \tau_{11} & \tau_{12} \\ \tau_{20} & \tau_{21} & \tau_{22} \end{bmatrix}$$

The matrix on the right-hand side is a variance-covariance matrix. The diagonal terms are the variances of the random effects, where τ_{00} represents the variance in school-mean achievement scores, τ_{11} represents the variance in gender contrasts across schools, and τ_{22} represents the variance in SES-Achievement slopes across schools.

The off-diagonal terms are covariances, where τ_{01} (and τ_{10}) capture the covariance between u_{0j} and u_{1j}, i.e., Cov(u_{0j}, u_{1j}); and τ_{02} (and τ_{20}) capture the covariance between u_{0j} and u_{2j}, i.e., Cov(u_{0j}, u_{2j}).

Among the parameters in the above matrix, the parameters of primary interest with respect to my analyses are the variance terms along the diagonal, particularly, τ_{11}, i.e., the variance in gender contrasts.

When predictors are entered into the level-2 equations, the diagonal terms represent the remaining variance in school-mean achievement, gender contrasts and SES-Achievement slopes, and the off-diagonal terms represent covariances conditional on the predictors in the level-2 equations.

Equations 2 to 4 are set up to explore variability in mathematics achievement scores across schools. In particular, equation 3 investigates school variability in gender differences for mathematics achievement.

In equation 2, γ_{00}indicates the grand mean of mathematics achievement scores for our population of schools and u_{0j} refers to the deviation of the mathematics score for school j from the expected grand mean. τ_{00} indicates the variation in school mean mathematics score around the grand mean.

In equation 3, γ_{10}refers to mean gender differences across schools. If the estimate of γ_{10} is significantly different from zero, it indicates that there are gender differences in mathematics performance on average. If the estimate of γ_{10} is negative, it indicates that on average males students outperform female students in mathematics achievement. If the estimate of γ_{10} is positive, it indicates that on average female students outperform male students in mathematics achievement. u_{1j} indicates the deviation of gender differences for school j from an expected value of grand mean gender differences. School variability in gender differences in mathematics performance is represented in the variance of u_{1j} (that is, τ_{11}). τ_{11}, one of the primary parameters of interest, indicates the school variability of gender differences in mathematics scores around the grand mean of gender differences. If τ_{11} is significantly different from zero, it suggests that the gender gap in mathematics performance varies across schools.

Results

TIMSS Regular 2015

Results of the Null Models

Descriptive statistics for the outcome variables are presented in Table 4.2. Tables 4.3 - 4.4 are descriptive statistics for each domain on the TIMSS assessment. Plausible values of mathematics performance for each domain were standardized so that 500 is the mean score and 100 is the standard deviation across all participating countries. On average, the weighted overall math achievement score for students in the USA was 518.99 (SD=80.82) with a range from 265.53 to 766.42.

Table 4.2. Descriptive Statistics for Five Plausible Values of Overall Mathematics Achievement TIMSS 2015 Regular (*N* = 8,352)

Plausible Values	Mean	Standard deviation	Min	Max
1st	518.77	82.36	248.22	762.23
2nd	518.99	83.51	213.18	778.08
3rd	519.63	83.95	203.74	796.56
4th	518.79	83.44	252.17	804.79
5th	518.76	83.98	223.99	800.6
Average	518.99	80.82	265.53	766.42

Table 4.3. Descriptive Statistics for Scores in the Content Domain (TIMSS 2015 Regular)

Variable name	Mean	Standard deviation	Min	Max
Algebra (averaged)	524.89	81.30	278.14	769.57
Number (averaged)	520.17	81.65	235.13	769.84
Geometry (averaged)	500.45	82.48	242.33	755.93
Data and Chance (averaged)	522.44	90.23	177.50	786.82

Table 4.4. Descriptive Statistics for Scores in the Cognitive Domain (TIMSS 2015 Regular)

Variable name	Mean	Standard deviation	Min	Max
Knowing (averaged)	528.61	84.00	276.56	779.26
Applying (averaged)	515.50	84.18	263.36	782.20
Reasoning (averaged)	514.33	77.85	257.72	766.86

Among areas in the content domain, the weighted average score for geometry was the lowest (*M* = 500.45). For the cognitive domain, the knowing area had the highest average score (*M* = 528.61).

Table 4.5. Parameter Estimates for the Null Model (TIMSS 2015 Regular)

Variables	Coefficient	*SE*	*t*	*p*
Fixed Effects				
Intercept (γ_{00})	520.56	3.49	149.21	<.001
Random Effects	Variance Component	*SD*	χ^2	*p*
School-level Effect (τ_{00})	2312.00	48.08	4906.56	<.001
Student-level Effect (σ^2)	4265.95	65.31		

$ICC = \frac{2312.00}{2312.00+4265.95} = 0.35$

Table 4.5 shows the results of the null model with overall mathematics performance scores as the outcome. The estimate of the grand mean of mathematics achievement scores for our population of schools is 520.56 (*p* <.001). Given the TIMSS Regular scale (*M* = 500, *SD* = 100), US students scored higher than the average of all participating students. The school level variability in mathematics proficiency is estimated at 2312.00 and there is significant variation in school mean mathematics proficiency, $\chi^2 = 4906.56$, *p* < .001. The student level variability is estimated at 4265.95. The ICC coefficient is 0.35, indicating that 35% of the total variance is between-schools [8]. The results of the null model also show an estimate of the reliability of school sample means. The reliability of the estimate is .96, indicating that the sample school means are quite reliable as indicators of the true school means.

Results for the Model for Question l

Descriptive statistics for predictor variables were calculated before conducting HLM analyses. Table 4.6 shows the minimum value, maximum value, weighted mean, and standard deviation for the student gender and SES variables. About half of the students in the sample were female and the mean SES of female students (M = 10.81, SD = 1.68) was similar to that of male students (M = 10.74, SD = 1.70). The results of the model for question 1 are presented in Table 4.7.

Table 4.6. Weighted Descriptive Statistics for Level 1 Variables (TIMSS 2015 Regular)

Variable	M	SD	Min	Max
Student Gender	0.51	0.51	0	1
Socio-economic Status (SES) of student	10.81	1.68	4.23	13.88

Table 4.7. Parameter Estimates for Question 1 (TIMSS 2015 Regular)

Variables	Coefficient	*SE*	*t*	*p*
Fixed Effects				
Intercept (γ_{00})	520.71	3.49	149.11	<.001
Within-School Effects				
Gender (γ_{10})	-4.20	2.09	-2.01	.053
SES (γ_{20})	11.55	0.77	15.08	<.001

[8] The ICC coefficients for each domain area were quite similar. In the content domain, ICC coefficient ranged from .32 (*number* area) to .34 (*algebra* area). In the cognitive domain, they ranged from .33 (*reasoning* area) to .35 (*knowing* area).

Random Effects	Variance Component	*SD*	χ^2	*P*
School-level Effect(τ_{00})	2327.65	48.25	5377.51	<.001
Gender contrast (τ_{11})	163.57	12.79	336.78	<.001
SES slope (τ_{22})	26.00	5.10	332.42	<.001
Student-level Effect (σ^2)	3899.12	62.44		

Student-level variance accounted for by L1 predictors = $\frac{(4265.95-3899.12)}{4265.95}$ = 0.09

Given that student-level variability is estimated at 4265.95 in the null model (see Table 4.5), the model for question 1 accounted for 9% of the student level variability by including student gender and SES. The grand mean of mathematics achievement scores for our population of schools is estimated at 520.71 (p <.001). Of the 201 schools, male outperformance was found in 122 schools, whereas female outperformance was found in 79 schools[9]. As student's SES increases by 1 unit, the average overall mathematics achievement score for the student increases by 11.55 points (p < .001). The mean gender difference in overall mathematics performance across schools is estimated at -4.20, indicating that male students outperform female students, though this difference is only marginally statistically significant (p = .053). A key aim of research question 1 is to investigate school variability in gender differences in mathematics performance and the primary coefficient of interest is the variance component. The variance component for the gender contrast is 163.57, which is statistically significant, $\chi^2 = 336.78$, p < .001 (see Table 4.7). This indicates that gender differences in mathematics performance vary across schools. Under the normality assumption, the 95% plausible value range for the gender gap slope is -4.2 ± 1.96*(12.79) = (-29.27, 20.87). Given that the standard deviation of

[9] Distribution of within-school gender gap in mathematics achievement across all schools is located in Appendix B (Figure B-1).

mathematics performance scores is 100, the range is almost half of a standard deviation. This range indicates that there is substantial variability in the size of the gender gap among schools.

Tables 4.8 and 4.9 show the gender gap for each area within the content and cognitive domains. In the content domain, the number area shows the largest gender gap (males outperform females). In fact, algebra is the only area in which female students outperform male students. The estimates of random effects for the gender contrast indicate that there is significant variability in the within-school gender gap (see Table 4.8). The geometry area has a smaller magnitude of between school variability compared to other areas.

Table 4.8. Gender Gap for areas in the Content Domain (TIMSS 2015 Regular)

Content Domain	Number		Algebra		Geometry		Data and Chance	
Fixed-effect	coefficient	(SE)	coefficient	(SE)	coefficient	(SE)	coefficient	(SE)
Gender	-12.64	(2.02)***	5.10	(2.17)*	-5.88	(2.57)*	-6.93	(2.30)**
Random Effects	variance component	(SD)	variance component	(SD)	variance component	(SD)	variance component	(SD)
Gender contrast	160.48	(12.67)***	166.31	(12.90)***	118.39	(10.88)***	171.46	(13.09)***

Note: N/S=Not significant; * *p*<.05; ** *p*<.01; ****p*<.001.

Table 4.9. Gender Gap for areas in the Cognitive Domain (TIMSS 2015 Regular)

Cognitive Domain	Knowing		Applying		Reasoning	
Fixed-effect	coefficient	(SE)	coefficient	(SE)	coefficient	(SE)
Gender	-1.19	(2.70)	-5.73	(2.17)*	-5.16	(2.64)
-						
Random Effects	variance component	(SD)	variance component	(SD)	variance component	(SD)
Gender contrast	187.82	(13.70)***	182.75	(13.52)***	166.31	(13.13)***

Note: N/S=Not significant; * *p*<.05; ** *p*<.01; ****p*<.001.

Assuming normality of the distribution of gender gaps across schools, the plausible value range for the gender gap can be constructed for each area. For the number area, the 95% plausible value range for the gender gap slope is -12.64 ± 1.96*(12.67) = (-37.47, 12.69). For the algebra area, the 95% plausible value range for the gender gap slope is 5.10 ± 1.96*(12.90) = (-20.18, 30.89). For the geometry area, the 95% plausible value range for the gender gap slope is -5.88 ± 1.96*(10.88) = (-27.20, 15.89). For the data and chance area, the 95% plausible value range for the gender gap slope is -6.93 ± 1.96*(13.09) = (-32.60, 19.26). Given that the standard deviation of mathematics performance scores for each area is 100, the range for each area is almost half of the standard deviation, indicating that there is large variability in the size of the gender gap in each content area among schools. For instance, even for the number area, which shows the largest gender gap favoring male students, there are some schools where the gender gap is positive, indicating that female students outperformed male students. In the cognitive domain, only the applying area showed a significant gender gap in mathematics performance, and it favored male students (see Table 4.9). The estimates of random effects for the gender contrast indicate that there is significant variability in the within-school gender gap for all the areas. Similar to the content domain, the range of plausible values can be constructed for each area under the normality assumption. For the knowing area, the 95% plausible value range for the gender gap slope is -1.19 ± 1.96*(13.70) = (-28.05, 26.22). For the applying area, the 95% plausible value range for the gender gap slope is -5.73 ± 1.96*(13.52) = (-32.22, 21.31). For the reasoning area, the 95% plausible value range for the gender gap slope is -5.16 ± 1.96*(13.13) = (-30.89, 21.10). Given that the standard deviation of mathematics performance scores is 100, the range for each area is almost half of the standard deviation, indicating that there is large variability in the size of the gender gap in each cognitive domain area among schools.

TIMSS Advanced 2015

Results of the Null Models

Descriptive statistics for the outcome variables and the weighted overall mean of each domain are presented in Tables 4.10 - 4.12. On average, the weighted overall math achievement score for advanced students in the USA was 494.10 (*SD* = 93.81) with a minimum score of 195.15 and a maximum score of 737.26.

Table 4.10. Descriptive Statistics for Five Plausible Values TIMSS 2015 Advanced (N = 2,274)

Plausible Values	Mean	Standard deviation	Min	Max
1st	493.12	99.71	141.82	788.23
2nd	494.31	97.68	143.37	767.40
3rd	493.59	97.07	157.38	780.61
4th	495.09	96.21	141.78	783.07
5th	494.38	97.13	193.42	772.42
Average	494.10	93.81	195.15	737.26

Table 4.11. Descriptive Statistics for Content Domain Scores (TIMSS 2015 Advanced)

Variable name	Mean	Standard deviation	Min	Max
Algebra (averaged)	486.89	92.86	170.99	742.00
Calculus (averaged)	514.87	103.18	150.57	802.57
Geometry (averaged)	463.52	97.71	56.82	750.17

Table 4.12. Descriptive Statistics for Content Domain Scores (TIMSS 2015 Advanced)

Variable name	Mean	Standard deviation	Min	Max
Knowing (averaged)	497.66	100.89	113.75	783.04
Applying (averaged)	489.36	93.70	164.00	727.25
Reasoning (averaged)	493.57	94.82	185.59	753.70

In the content domain, the lowest weighted average score was in geometry (*M* = 463.52). In the cognitive domain, the knowing area had the highest average score (*M* = 497.66).

Table 4.13. Parameter Estimates for the Null Model (TIMSS 2015 Advanced)

Variables	Coefficient	*SE*	*t*	*p*
Fixed Effects				
Intercept (γ_{00})	473.62	6.00	78.89	<.001
Random Effects	Variance Component	*SD*	χ^2	*p*
School-level Effect (τ_{00})	5406.99	73.53	2176.36	<.001
Student-level Effect (σ^2)	5482.79	74.05		

$$ICC = \frac{5406.99}{5406.99+5482.79} = 0.50$$

Table 4.13 shows the results of the null model for overall mathematics performance scores. The grand mean of mathematics achievement scores for our population of schools is estimated at 473.62 (p <.001). The school-level variability in mathematics proficiency is estimated at 5406.99, and there is significant variation in school mean mathematics proficiency, $\chi^2 = 2176.36$, $p < .001$. The student level variability is estimated at 5482.79. The ICC coefficient is 0.50, indicating that 50% of the total variance is attributed to between-school variance. The results of the null model also show an estimate of the reliability of school sample means. The reliability of the estimate is .92, indicating that the sample means are quite reliable as indicators of the true school means.

Results for the Model for Question l

Table 4.14 shows the minimum value, maximum value, weighted mean, and standard deviation for student gender and SES. About half of the students in the sample were female and the mean SES of female students (*M* = 10.54, *SD* = 1.95) was similar to that of male students (*M* = 10.46, *SD* = 1.93).

Table 4.14. Weighted Descriptive Statistics for Level 1 Variables (TIMSS 2015 Advanced)

Variable	M	SD	Min	Max
Student Gender	0.49	0.50	0	1
Socio-economic Status (SES) of student	10.5	1.94	1.49	13.29

The results of the model for question 1 are presented in Table 4.15.

Table 4.15. Parameter Estimates for Question 1 (TIMSS 2015 Advanced)

Variables				
	Coefficient	*SE*	*t*	*p*
Fixed Effects				
Intercept (γ_{00})	473.02	6	78.82	<.001
Within-School Effects				
Gender (γ_{10})	-25.75	4.35	-5.92	<.001
SES (γ_{20})	3.91	1.18	3.32	0.002
Random Effects	Variance Component	*SD*	χ^2	*P*
School-level Effect (τ_{00})	5426.16	73.66	2330.41	<.001
Gender contrast (τ_{11})	1144.4	33.83	316.85	<.001
Student-level Effect (σ^2)	4994.85	70.67		

Student-level variance accounted for by L1 predictors = $\frac{(5482.79-4971.60)}{5482.79} = 0.09$

Given that student-level variability is estimated at 5482.79 in the null model (see Table 4.13), the model for question 1 accounted for 9% of the student level variance by including student gender and SES. The grand mean of mathematics achievement scores for our population of schools is estimated at 473.02 (*p* <.001). In addition, as student SES increases by 1 unit, their average overall mathematics achievement score increases by 3.91 points (*p* = .002). The mean gender difference in overall mathematics performance across schools is estimated at -25.75, indicating that male students significantly outperform females (p <.001).

The variance component for the gender contrast is 1144.40 and is statistically significant, $\chi^2 = 316.85$, p < .001. This indicates that gender differences in mathematics performance vary across schools. Under the normality assumption, the 95% plausible value range for the gender gap slope is -25.75± 1.96*(33.83) = (-92.05, 40.56). Given that the standard deviation for mathematics performance scores is 100, the range is about 1.3 times that of the standard deviation. This range indicates that schools vary significantly in the degree of gender differences on mathematics scores. The range also shows that there are schools in which female students outperformed male students.

Tables 4.16 and 4.17 show the gender gap for each domain area.

Table 4.16. Gender Gap for areas in the Content Domain (TIMSS 2015 Advanced)

Cognitive Domain	Algebra		Calculus		Geometry	
Fixed-effect	coefficient	(SE)	coefficient	(SE)	coefficient	(SE)
Gender	-17.56	(5.08)***	-18.81	(5.73)*	-33.65	(5.07)*
-						
Random Effects	variance component	(SD)	variance component	(SD)	variance component	(SD)
Gender contrast	1597.90	(39.97)***	1256.70	(35.45)***	844.32	(29.06)***

Note: N/S=Not significant; * p<.05; ** p<.01; ***p<.001.

Table 4.17. Gender Gap for areas in the Cognitive Domain (TIMSS 2015 Advanced)

Cognitive Domain	Knowing		Applying		Reasoning	
Fixed-effect	coefficient	(SE)	coefficient	(SE)	coefficient	(SE)
Gender	-21.96	(6.60)**	-16.16	(4.97)**	-27.60	(5.60)***
-						
Random Effects	variance component	(SD)	variance component	(SD)	variance component	(SD)
Gender contrast	1485.32	(38.54)***	1164.46	(34.12)***	1116.72	(33.42)***

Note: N/S=Not significant; * p<.05; ** p<.01; ***p<.001.

All areas in the content domain show a significant gender gap, with male students outperforming females (on average). In terms of the magnitude of gender gaps, geometry has the largest gender gap. The estimates of random effects for the slope of the gender variable indicate that there is significant variability in gender gaps across schools (see Table 4.16). The geometry area has a smaller magnitude of between school variability compared to other areas. Assuming normality of the distribution of gender gaps, the plausible value range for gender gaps can be calculated for each area. For the algebra area, the 95% plausible value range for the gender gap slope is $-17.56 \pm 1.96*(39.97) = (-95.91, 62.38)$. For the calculus area, the 95% plausible value range for the gender gap slope is $-18.81 \pm 1.96*(35.45) = (-88.29, 52.09)$. Given the standard deviation of mathematics performance (100), the ranges for the algebra and calculus areas are almost 1.5 times the standard deviation, indicating that there is substantially large variability in the size of the gender gap for those areas. For the geometry area, the 95% plausible value range for the gender gap slope is $-33.65 \pm 1.96*(29.06) = (-90.60, 24.46)$. The range is more than one standard deviation, indicating that there is substantially large variability in the size of the gender gap in the geometry area among schools.

All areas in the cognitive domain also show a significant gender gap favoring male students. The reasoning area has a gender gap with the largest magnitude. The estimates of the random effects for the slope of the gender variable indicate that there is significant within-school gender gap variability for all the areas.

The range of plausible values for each area shows that there is substantial variability in the magnitude of the gender gap for each area. For the knowing area, the 95% plausible value range for the gender gap slope is $-21.96 \pm 1.96*(38.54) = (-97.50, 55.12)$ under the normality assumption. Given that TIMSS scores are scaled to have a standard deviation of 100 points, the range is almost 1.5 times that of the standard deviation. For the applying area, the 95% plausible value range for the gender gap slope is -16.16 ± 1.96*(34.12) = (-83.04, 52.09). For the reasoning area, the 95% plausible value range for the gender gap slope is -27.60 ± 1.96*(33.42) = (-93.10, 39.23). The ranges for the applying and reasoning areas are almost 1.3 times the standard deviation of TIMSS scores.

Discussion and Implications

The findings of large school-to-school differences in student achievement in the null models suggest that there is considerable similarity among students in terms of achievement in the same school. For 8th grade students (TIMSS 2015 Regular), about 35% of the variance in students' mathematics performance is between schools. Twelfth grade students who take advanced mathematics classes show even greater similarity among students– about 50% of the variance in mathematics performance is attributed to between schools. These results indicate that a shared school environment may be substantially related to students' performance; thus, it is important to investigate the relationship between school environment and students' mathematic performance.

When it comes to the hypothesized gender gap in mathematics performance, the results show that male students perform better than female students, on average, at both 8th and 12th grades. In addition, the current study found that the magnitude of the gender gap for 12th grade students was greater than that of 8th grade students. Indeed, the difference between males and females in 8th grade was marginally significant, whereas 12th graders taking advanced mathematics classes show a significant and wide gender gap in mathematics performance. These results are consistent with previous research that has found that the gender gap in mathematics, favoring male students, tends to be larger at higher grades and among high-achieving students (e.g., Hyde, Fennema, & Lamon, 1990).

The results also indicate that the pattern of gender differences in mathematics performance scores differs across areas of mathematics. For example, in the content domain, *algebra* was the only sub-domain in which females outperform males, though this was only for 8^{th} grade students. This is consistent with previous research finding that female students are usually better at solving algebra items, likely because it requires memorizing equations and formulas (Gallagher, 1998). In contrast, the *number* sub-domain had the largest gender gap favoring male students. Again, this is consistent with prior research using data from PISA (Programme for International Student Assessment; Liu & Wilson, 2009).

In addition, the geometry sub-domain showed that males outperformed females, on average. Previous research reports that male students perform better at geometry, which requires visual-spatial skills (Carlton & Harris, 1989; O'Neill, Wild, & McPeek, 1989). Using PISA data, Liu and Wilson (2009) also found that geometry showed a gender gap favoring male students. The purpose of the PISA mathematics assessment is different from that of TIMSS mathematics. While the TIMSS assessment is based on mathematics curricula, the PISA assessment looks at students' ability to solve mathematical problems in various real world situations. Nevertheless, even though the purposes of the assessments are different, the pattern of males outperforming females in the content domain is very similar.

In the cognitive domain for 8^{th} grade students, all sub-domains showed male student outperformed female students. The *knowing* area had the smallest gender gap, which is consistent with prior work that has found that male students are better at solving unconventional items which require applying and reasoning rather than applying knowledge (Gallagher & De Lisi, 1994).

The main research question of the current study is to investigate gender gap variability in mathematics performance across schools. Thus, the variance of gender gaps is the main parameter of interest. These results show that there is a considerable variability between schools in the size of the gender gap in mathematics performance. For 8th grade students, the range of the gender gap is about half of the standard deviation. More compellingly, in some schools, females outperform males in mathematics scores. Variability across schools in mathematics performance assessment scores between genders gets even greater at 12th grade high achieving students. Indeed, the range of the gender gap is more than a standard deviation at 12th grade for the TIMSS 2015 Advanced. The current results suggest a need to investigate factors that are related to school variability in the size of the gender gap in mathematics achievement. Chapter 5 explores this question. Specifically, Chapter 5 focuses on overall mathematics scores (five plausible values) as the outcome (rather than the separate mathematics sub-scores) and investigates the factors that are related to the large variability in the gender gap across schools[10].

[10] The correlations between mathematics scores for each area was above .95 so Chapter 5 focuses on overall mathematics scores, instead of using scores for each area.

Chapter 5. Research Question 2: Results of School-Level HLM

<u>*Research Question 2:*</u> What school-level factors are associated with school variability in gender differences in mathematics achievement?

Upon discovery that gender differences in mathematics performance vary across schools, it is crucial to investigate which school factors are associated with school variability. For question 2, level II (school-level) models were specified using school-level variables in order to investigate which school variables are related to school variability in the mathematics gender gap[11].

TIMSS 2015 Regular

Data

The same data used in question 1 were used in question 2.

<u>*Variables of Interest*</u>

Table 5.1 shows a brief description of the student and school-level variables used in the models.

Table 5.1. Student and School Variables (TIMSS 2015 Regular)

Variable Name	Description
Outcome	
Math PVs	Mathematics plausible values for the TIMSS 2015 8th grade students
Student-level	
GEN	Gender of student
SES	Socio-economic status of students
School-level	
BCBGMRS	Instruction affected by mathematics resources shortage
MEAN SES	Socio-economic Status (SES) of schools
TCONF	Teacher confidence in teaching mathematics

[11] In this chapter, the outcome variable is limited to the overall mathematics score, represented by five plausible values. The mathematics scores for each area (content and cognitive domain) were not used.

TCOM	Communication among teachers
TSUP	Teacher support of student participation
SHORTT	Shortage of teachers

Additionally, histograms showing the distributions of each variable and scatterplots between variables are presented in Appendix B (Figure B-1 – B-21).

Student-level Variables

The same student variables used for question 1 were used in question 2.

- *Gender of student (GEN)*: Gender of student is a main variable of interest. In the models, male students were coded as 0 and female students were coded as 1.
- *Socio-economic status (SES) of students*: The Home Educational Resources scale, provided by the TIMSS administration, was used as a proxy of individual student SES. It is a derived variable based on the three components: 1) number of books in the home, 2) number of home study supports (i.e., whether students have internet connection and/or their own room), and 3) highest level of education of either parent.
- *Outcome variables*: The main outcome variable for the study is overall mathematics performance of students, represented by a set of five plausible values.

School-level Variables

- *Instruction affected by mathematics resources shortage (BCBGMRS)*: This is a variable that was created and provided by the TIMSS administration. School principals were asked to answer how much of their school's capacity to provide instruction is affected by a shortage or inadequacy of resources for mathematics instruction. They were asked to answer using a four-scale response (not at all, a little, some, and a lot) to the following items.

a) Teachers with a specialization in mathematics

b) Computer software/applications for mathematics instruction

c) Library resources relevant to mathematics instruction

d) Calculators for mathematics instruction

e) Concrete objects or materials to help students understand quantities or procedures

This variable was scored using the IRT partial credit model developed by the TIMSS administration and standardized so that the mean is 10 and the standard deviation is 2 across all TIMSS participating countries. A higher score means that mathematics instruction is less affected by resource shortages.

- *Socio-economic Status (SES) of schools (MEAN SES):* This variable is aggregated from the students' home educational resources scale (student SES) within each school to represent socio-economic status (SES) of schools. This measure is used as a proxy of school SES since students' aggregated responses can represent the "shared perception of the environment" (Lüdtke, Robitzsch, Trautwein, & Kunte, 2009).
- *Teacher Confidence in Teaching Mathematics (TCONF):* This is a derived variable from teacher responses that was created for the purpose of this study. I selected items that capture *teacher confidence in teaching mathematics* and conducted exploratory IRT[12]. Items with high factor loadings (higher than .6)

[12] The R output for the exploratory IRT is in Appendix A.

were selected to estimate *teacher confidence in teaching mathematics* and scored using an IRT graded response model (GRM). Teachers were asked to characterize their confidence in teaching class by selecting from four categories (very high, high, medium, and low) for each of the following items:

a) Inspiring students to learn mathematics
b) Showing students a variety of problem solving strategies
c) Providing challenging tasks for the highest achieving students
d) Adapting my teaching to engage students' interest
e) Helping students appreciate the value of learning mathematics
f) Assessing student comprehension of mathematics
g) Improving the understanding of struggling students
h) Making mathematics relevant to students
i) Developing students' higher-order thinking skills

This variable was standardized so that the mean is 10 and the standard deviation is 2 for all sampled teachers in the United States. Higher scores indicate higher teacher confidence in teaching mathematics. The estimated reliability (Cronbach's alpha) for the *teacher confidence in teaching mathematics* variable was .92.

- *Communication among Teachers (TCOM)*: This is a derived variable from teacher responses that was created for the purpose of the study. I selected items that capture *communication among teachers* and conducted exploratory IRT[13]. Items with high factor loadings (higher than .6) were selected to estimate *communication among teachers* and scored using an IRT GRM. Teachers were

[13] The R output for the exploratory IRT is in Appendix A.

asked how often they have the following types of interaction with other teachers by selecting from four categories (very often, often, sometimes, and never or almost never).

a) Discuss how to teach a particular topic

b) Collaborate in planning and preparing instructional materials

c) Share what I have learned about my teaching experiences

d) Work together to try out new ideas

e) Work as a group on implementing the curriculum

This variable score was standardized so that the mean is 10 and the standard deviation is 2 for all sampled teachers in the United States. Higher scores indicate that a teacher interacts more frequently with other teachers. The estimated reliability (Cronbach's alpha) for the *communication among teachers* variable was .92.

- *Teacher support of student participation (TSUP):* This is a derived variable from teacher responses that was created for the purpose of this study. Though TIMSS administered items concerning teacher instructional practices, TIMSS did not provide any latent variables for those items. Thus, I selected items that capture *teacher support of student participation* and conducted exploratory IRT[14]. Items with high factor loadings (higher than .6) were selected to estimate *teacher support of student participation* and scored using an IRT GRM. Teachers were asked how often they did the following while teaching the class by selecting from four answer categories (very often, often, sometimes, and never or almost never):

[14] The R output for the exploratory IRT is in Appendix A.

a) Ask students to explain their answers

b) Encourage classroom discussions among students

c) Ask students to decide their own problem solving procedures

d) Encourage students to express their ideas in class

This variable was standardized so that the mean is 10 and the standard deviation is 2 for all sampled teachers in the United States. Higher scores indicate more frequent teacher support for student participation. The estimated reliability (Cronbach's alpha) for the *teacher support of student participation* variable was .76.

- *Shortage of teachers (SHORTT):* This is a derived binary variable from school principal responses that was created for the purpose of the study. School principals were asked how difficult it has been this school year to fill 8th grade mathematics and science teaching vacancies. If they said that a) there were no vacancies for mathematics and science subjects or b) it is easy to fill vacancies for mathematics or sciences, their responses were coded as 0 (no shortage of teachers). If a school principal answered that a) it is somewhat difficult to fill or b) it is very difficult to fill, they were coded as 1 (schools with a shortage of teachers).

Model Specification

At the student-level (level I), individual students' mathematics achievement scores are modeled as a linear relationship of students' gender and students' SES, as was posed in question 1. The student weight variable (STUDWGT) was used for the level 1 model.

Level I: Student-level model

$$Y_{ij} = \beta_{0j} + \beta_{1j} Gender_{ij} + \beta_{2j} SES_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

β_{0j} indicates the mean mathematics score for the 8th grade students in school j. The level I covariates (SES_{ij} and $Gender_{ij}$) are expressed as deviations from the school means using group-mean centering in the level I model.

To investigate the school-level variables that are associated with the within-school gender gap in mathematics achievement, school-level variables were included in the level 2 (school-level) model. The school weight variable was used for the level 2 model.

School-level models are as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\boldsymbol{BCBGMRS}_{1j} + \gamma_{02}\boldsymbol{MEAN\ SES}_{2j} + \gamma_{03}\boldsymbol{TCONF}_{3j} + \gamma_{04}\boldsymbol{TCOM}_{4j} + \gamma_{05}\boldsymbol{TSUP}_{5j} + \gamma_{06}\boldsymbol{SHORTT}_{6j} + u_{0j},$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\boldsymbol{BCBGMRS}_{1j} + \gamma_{12}\boldsymbol{MEAN\ SES}_{2j} + \gamma_{13}\boldsymbol{TCONF}_{3j} + \gamma_{14}\boldsymbol{TCOM}_{4j} + \gamma_{15}\boldsymbol{TSUP}_{5j} + \gamma_{16}\boldsymbol{SHORTT}_{6j} + u_{1j},$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}\boldsymbol{BCBGMRS}_{1j} + \gamma_{22}\boldsymbol{MEAN\ SES}_{2j} + \gamma_{23}\boldsymbol{TCONF}_{3j} + \gamma_{24}\boldsymbol{TCOM}_{4j} + \gamma_{25}\boldsymbol{TSUP}_{5j} + \gamma_{26}\boldsymbol{SHORTT}_{6j} + u_{2j},$$

$$\mathrm{Var}\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{bmatrix} = \begin{bmatrix} \tau_{00} & \tau_{01} & \tau_{02} \\ \tau_{10} & \tau_{11} & \tau_{12} \\ \tau_{20} & \tau_{21} & \tau_{22} \end{bmatrix}$$

The matrix on the right-hand side is a variance-covariance matrix. The diagonal terms are the variances of the random effects, where τ_{00} represents the variance in school-mean achievement scores, τ_{11} represents the variance in gender contrasts across schools, and τ_{22} represents the variance in SES-Achievement slopes across schools.

The off-diagonal terms are covariances, where τ_{01} (and τ_{10}) capture the covariance between u_{0j} and u_{1j}, i.e., Cov(u_{0j}, u_{1j}); and τ_{02} (and τ_{20}) capture the covariance between u_{0j} and u_{2j}, i.e., Cov(u_{0j}, u_{2j}).

Among the parameters in the above matrix, the parameters of primary interest with respect to my analyses are the variance terms along the diagonal, particularly, τ_{11}, i.e., the variance in gender contrasts.

When predictors are entered into the level-2 equations, the diagonal terms represent the remaining variance in school-mean achievement, gender contrasts and SES-Achievement slopes, and the off-diagonal terms represent covariances conditional on the predictors in the level-2 equations.

In the three equations above, school-level variables were included to explain school variability in gender gaps, whereas no school-level predictors were included in the level II models in the previous chapter (Chapter 4). Predictors in the level II model are expressed as deviations from the grand mean (grand-mean centering). The coefficients γ_{1m} (m=1, …, 6) represent the relationships between the within-school gender gap and the school-level factors and are of primary interest for research question 2.

Results

<u>*Descriptive Statistics for the Variables*</u>

Table 5.2 shows the minimum value, maximum value, weighted mean, and standard deviation of each variable included in the models.

The *instruction affected by mathematics resources shortage* variable was scaled such that its mean is 10 and the standard deviation is 2 across all TIMSS participating countries. The US school sample used in the study had a weighted mean of 10.78 and a standard deviation of 1.66.

The socio-economic status (SES) of schools variable was aggregated from the student SES variable. The mean of US school sample has a weighted mean of 11.16 and a standard deviation of 0.95. The *teacher confidence in teaching mathematics*, *communication among teachers*, and *teacher support of student participation* variables have weighted means close to 10 and standard deviations close to 2. The *Shortage of teachers variable* has a weighted mean of 0.24 and a standard deviation of 0.43 for the all sampled US schools.

Table 5.2. Descriptive Statistics of the Variables (TIMSS 2015 Regular)

Variable	M	SD	Min	Max
Instruction affected by mathematics resources shortage	10.78	1.66	4.05	14.56
Socio-economic Status (SES) of schools	11.16	0.95	8.76	13.25
Teacher confidence in teaching mathematics	10.00	1.78	5.7	13.61
Communication among teachers	9.28	1.83	5.37	13.09
Teacher support of student participation	10.42	1.75	5.88	12.5
Shortage of teachers	0.24	0.43	0	1

<u>Correlation among Variables</u>

Bivariate correlation analyses were conducted to explore the relations among variables. The analysis with the two variables in the level 1 model – student gender and student SES – indicates that the two variables are not correlated (r = .02, p >.05).

Unweighted bivariate relationships between the level 2 variables were estimated. Table 5.3 presents the correlation matrix for these variables. The correlation coefficients of the level 2 variables ranged from -.13 (between *instruction affected by mathematics resources shortage* and *shortage of teachers*) to .41 (between *teacher confidence in teaching mathematics* and *teacher support of student participation*). High correlation coefficients might cause multicollinearity issues in the HLM analyses, but the correlation analysis with the level 2 variables shows a relatively weak relationship among the variables. *Instruction affected by mathematics resources shortage* has a positive correlation with *school SES* (r = .21) and a negative correlation with *shortage of teachers* (r = .13). *Teacher confidence in teaching mathematics, communication among teachers*, and *teacher support of student participation* have positive bivariate correlations with each other. Scatterplots showing the relationships between each pair of variables are presented in Figure 5.1.

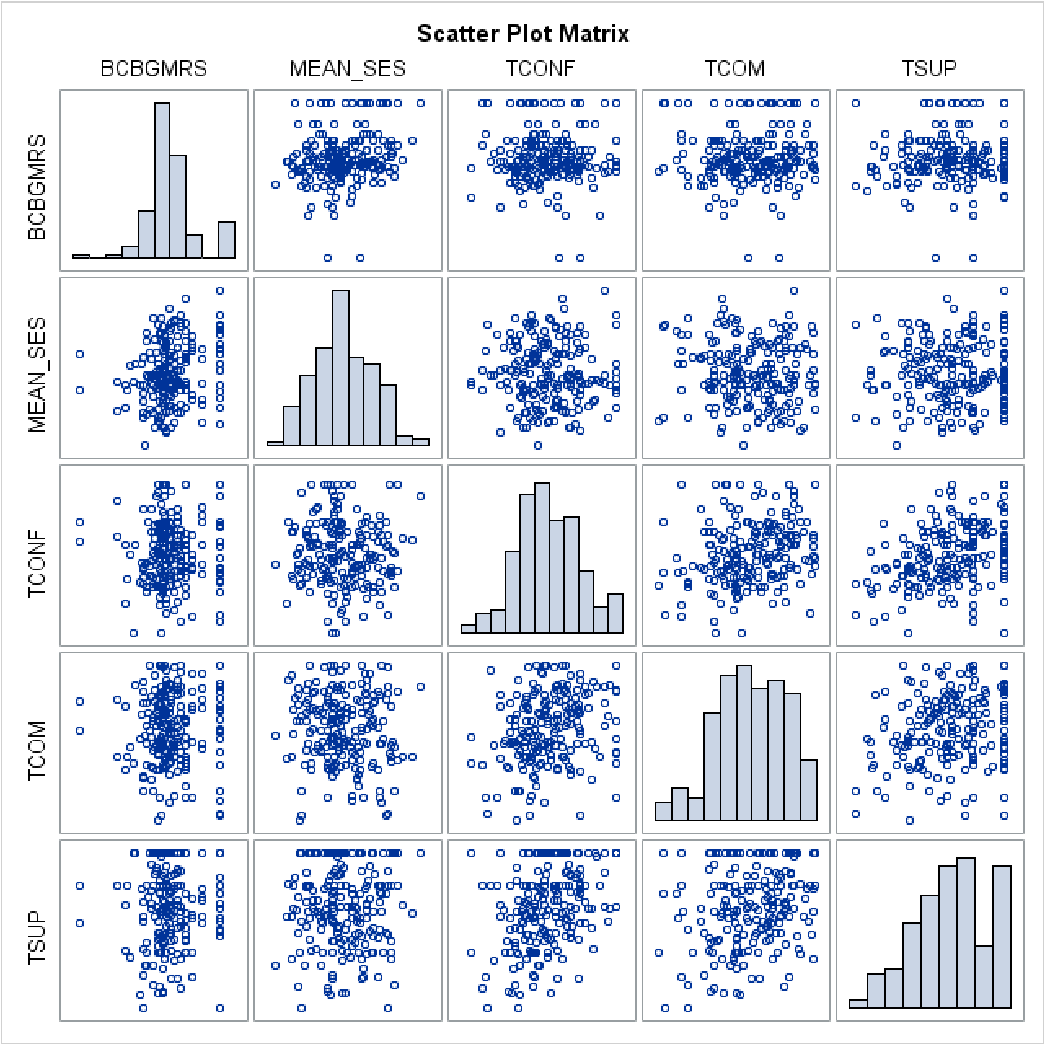


Figure 5.1. Scatter Plot Matrix for variables (TIMSS 2015 Regular)

Table 5.3. Correlation among Variables (TIMSS 2015 Regular)

Variable	*Instruction affected by mathematics resources shortage*	*Socio-economic Status (SES) of schools*	*Teacher confidence in teaching mathematics*	*Communication among Teachers*	*Teacher support of student participation*	*Shortage of teachers*
Instruction affected by mathematics resources shortage	1.00	0.21**	0.01	0.00	0.02	-0.13†
Socio-economic Status (SES) of schools		1.00	-0.03	-0.09	0.01	-0.13
Teacher confidence in teaching mathematics			1.00	0.25***	0.41***	-0.01
Communication among Teachers				1.00	0.22**	0.10
Teacher support of student participation					1.00	-0.01
Shortage of teachers						1.00

Note: †p<.10. *p<.05. **p<.01. ***p<.001.

Results for the Model for Question 2

The results of the model for question 2 are presented in Table 5.4.

Table 5.4. Results of HLM analysis (TIMSS 2015 Regular)

Variables	Coefficient	*SE*	*t*	*p*
Fixed Effects				
mathematics achievement				
Intercept	512.16	2.53	202.76	<.001
BCBGMRS	3.49	1.46	2.39	0.02
MEAN SES	37.28	2.58	14.48	<.001
TCONF	1.34	1.53	0.87	0.38
TCOM	3.2	1.37	2.34	0.02
TSUP	-0.13	1.49	-0.09	0.93
SHORTT	-15.89	5.56	-2.86	0.005
Gender				
Intercept	-4.35	2.01	-2.17	0.036
BCBGMRS	-3.31	1.05	-3.16	0.002
MEAN SES	3.5	1.95	1.79	0.075
TCONF	-2.43	1.05	-2.32	0.022
TCOM	1.97	1.02	1.93	0.055
TSUP	3.1	1.18	2.62	0.011
SHORTT	5.48	4.51	1.22	0.23
Student SES				
Intercept	10.45	0.7	14.87	<.001
BCBGMRS	-0.34	0.36	-0.93	0.36
MEAN SES	4.62	0.74	6.25	<.001
TCONF	-0.89	0.4	-2.21	0.03
TCOM	0.003	0.36	-0.93	0.36
TSUP	0.67	0.4	1.69	0.1
SHORTT	2.17	1.48	1.47	0.15

Random Effects	Variance Component	*SD*		*P*
Mathematics achievement	908.83	30.15	2343.83	<.001
Gender contrast	88.96	9.43	298.62	<.001
Student SES slope	9.41	3.07	254.27	0.003
Student-level Effect	3894.76	62.41		

Gender contrast variance accounted for by L2 predictors = $\frac{(163.57-88.96)}{163.57} = 0.46$

The variance of the gender contrast for this model is 88.96, whereas the variance of the gender contrast is 163.57 in the model for question 1 (the random-coefficients regression model), which only includes level 1 predictors (student SES and gender). Approximately 46% of the gender contrast variance is accounted for by the level 2 predictors in the model for question 2. The grand mean of overall mathematics achievement scores for our population of schools is estimated at 512.16 (p <.001). All variables included in the model are significantly related to the school mean mathematics achievement scores except for *teacher confidence in teaching mathematics* (TCONF) and *teacher support of student participation* (TSUP). For instance, as *instruction affected by mathematics resources shortage* (BCBGMRS) increases by 1 unit, the school-mean overall mathematics achievement scores increase by 3.49 points (p = .02), holding constant other variables. Note that a higher score means that mathematics instruction is less affected by resource shortages. As school SES (MEAN SES) increases by 1 unit, school-mean overall mathematics achievement scores increase by 37.28 points (p <.001), holding constant other variables in the model. *Shortage of teachers* (SHORTT) is negatively related to mean mathematics achievement, whereas *communication among teachers* (TCOM) is positively related to mean mathematics achievement. The school-mean overall mathematics achievement score for schools with teacher shortage is 15.89 points (p = .005) lower than that of schools without a shortage, holding constant other variables in the model. As *communication among teachers* (TCOM) increases by 1 unit, school-mean overall mathematics achievement scores increase by 3.2 points (p = .02), holding constant other variables.

Several variables are associated with the gender contrast, which is of primary interest in research question 2. A negative value for the gender contrast indicates male outperformance, while a positive value indicates female outperformance. The intercept for the gender contrast is -

4.35. This means that the expected gender gap for a school whose values for the predictors (BCBGMRS, MEAN SES, TCONF, TCOM, TSUP, and SHORTT) are equal to the grand means of those predictors is -4.35, indicating male outperformance. *Instruction affected by mathematics resources shortage* (BCBGMRS) and *teacher confidence in teaching mathematics* (TCONF) are negatively related to the magnitude of the student gender contrast, while *teacher support of student participation* (TSUP) is positively related to the magnitude of the student gender contrast. *Communication among teachers* (TCOM) is also positively related to the magnitude of the student gender contrasts, although the relationship is marginally significant at the .05 alpha level (p = .055).

As *instruction affected by mathematics resources shortage* (BCBGMRS) increases by 1 unit, the gender contrast decreases by 3.31 points, holding constant other variables. As *teacher confidence in teaching mathematics* (TCONF) increases by 1 unit, the gender contrast decreases by 2.43 points, holding constant other variables. For *communication among teach*ers (TCOM), a 1 unit increase in TCOM is associated with a 1.97 point increase in the magnitude of gender contrasts holding other variables constant, a marginally significant effect (p = .055). A 1 unit increase in *teacher support of student participation* (TSUP) is associated with a 3.1 point increase in the magnitude of gender contrasts, holding other variables constant.

To facilitate the interpretation, consider two schools whose values for other predictors (MEAN SES, TCONF, TCOM, TSUP, and SHORTT) are equal to the grand means of those predictors except for the *instruction affected by mathematics resources shortage* variable. One school (School A) has a value of *instruction affected by mathematics resources shortage* variable that is one standard deviation (1.78) above the grand mean for *instruction affected by mathematics resources shortage* variable. The other school (School B) has a value of *instruction* *affected by mathematics resources shortage* variable that is one standard deviation (1.66) below the grand mean for *instruction affected by mathematics resources shortage* variable. For School A, the expected gender gap would be -4.35 – 3.31* (1.66) = -9.84, indicating that male students are expected to outperform female students by 9.84 points in School A. For School B, the expected gender gap would be -4.35 - 3.31* (-1.66) = 1.14, indicating that female students are expected to outperform male students by 1.14 points in School B. Considering that the student-level random effect is 65.31 in the null model (see Chapter 4), the 10.98 (=1.14 – (-9.84)) points difference is about 17% of the standard deviation of the level 1 random effect. A 95% confidence interval around the point estimate of *instruction affected by mathematics resources shortage* is from -5.37 to -1.25, indicating that we are 95% confident that our interval contains or captures the true value of the coefficient for the *instruction affected by mathematics resources shortage* predictor variable. This means that the true value of the coefficient may be as small as -5.37 or as large as -1.25. Note that the range of plausible values for the true value of the coefficient of interest spans only negative values.

With respect to *teacher confidence in teaching mathematics (TCONF*), consider two schools whose values for other predictors (BCBGMRS, MEAN SES, TCOM, TSUP, and SHORTT) are equal to the grand means of those predictors except for the *teacher confidence in teaching mathematics* variable. One school (School C) has a value of *teacher confidence in teaching mathematics* variable that is one standard deviation (1.78) above the grand mean for *teacher confidence in teaching mathematics* variable. The other school (School D) has a value of *teacher confidence in teaching mathematics* variable that is one standard deviation (1.78) below the grand mean for *teacher confidence in teaching mathematics* variable. For School C, the expected gender gap would be -4.35 – 2.43* (1.78) = -8.68, indicating that male students are expected to outperform female students by 8.68 points in School C. For School D, the expected gender gap would be -4.35 – 2.43* (-1.78) = -0.02, indicating that male students are expected to outperform female students by 0.02 points in School D.

For *communication among teachers* variable, consider two schools whose values for other predictors (BCBGMRS, MEAN SES, TCONF, TSUP, and SHORTT) are equal to the grand means of those predictors except for the *communication among teachers*. One school (School E) has a value of *communication among teachers* variable that is one standard deviation (1.83) above the grand mean for *communication among teachers* variable. The other school (School F) has a value of *communication among teachers* variable that is one standard deviation (1.83) below the grand mean for *communication among teachers* variable. For School E, the expected gender gap would be -4.35 + 1.97* (1.83) = -.74, indicating that male students are expected to outperform female students by 0.74 points in School E. For School F, the expected gender gap would be -4.35 + 1.97* (-1.83) = -7.96, indicating that male students are expected to outperform female students by 7.96 points in School F. Considering that the student-level random effect is 65.31 in the null model (see Chapter 4), the 7.22 (=-.74 – (-7.96)) points difference is about 11% of the standard deviation of the level 1 random effect. A 95% confidence interval around the point estimate of *communication among teachers* ranges from -0.03 to 3.97, indicating that we are 95% confident that our interval contains or captures the true value of the coefficient for the *communication among teachers* variable. Note that the range of plausible values for the true value of the coefficient of interest includes a negative lower bound (with a small magnitude), though most of the range spans positive values.

With respect to *teacher support of student participation* variable, consider two schools whose values for other predictors (BCBGMRS, MEAN SES, TCONF, TCOM, and SHORTT) are equal to the grand means of those predictors except for the *teacher support of student participation*. One school (School X) has a value of *teacher support of student participation* variable that is one standard deviation (1.75) above the grand mean for *teacher support of student participation* variable. The other school (School Y) has a value of *teacher support of student participation* variable that is one standard deviation (1.75) below the grand mean for *teacher support of student participation* variable. For School X, the expected gender gap would be -4.35 + 3.1* (1.75) = 1.08, indicating that female students are expected to outperform male students by 1.08 points in School X. For School Y, the expected gender gap would be -4.35 + 3.1* (-1.75) = -10.85, indicating that male students are expected to outperform female students by 10.85 points in School Y. Considering that the student-level random effect is 65.31 in the null model (see Chapter 4), the 11.93 (=1.08 – (-10.85)) points difference is about 18% of the standard deviation of the level 1 random effect. A 95% confidence interval around the point estimate of *teacher support of student participation* ranges from 0.79 to 5.41, indicating that we are 95% confident that our interval contains or captures the true value of the coefficient for the *teacher support of student participation* variable. Note that the range of plausible values for the true value of the coefficient of interest spans only positive values.

Regarding the slope for student SES, school mean SES shows a positive interaction effect whereas *teacher confidence in teaching mathematics* (TCONF) shows a negative interaction effect. As school SES increases by 1 unit, the effect of student SES on student mathematics achievement scores increases by 4.62. The positive interaction indicates that school SES has an effect on the mathematics achievement above and beyond student SES. As *teacher confidence* increases by 1 unit, the effect of student SES on student mathematics achievement scores decreases by 0.89.

TIMSS 2015 Advanced

Data

The same data used in question 1 was used in question 2.

Variables of Interest

Table 5.5 shows a brief description of the student and school-level variables used in the model.

Table 5.5. Description of Variables in the model (TIMSS 2015 Advanced)

Variable Name	Description
Outcome	
Math PVs	Mathematics plausible values for the TIMSS 2015 12th grade high achieving students
Student-level	
GEN	Gender of student
SES	Socio-economic status of students
School-level	
MTBGSOS	Safe and orderly school
MEAN SES	Socio-economic Status (SES) of schools
TCOM	Communication among teachers
TSUP	Teacher support of student participation
SHORTT	Shortage of teachers

Additionally, histograms showing the distributions of each variable and scatterplots between variables are presented in Appendix C (Figure C-1 – C-15).

Student-Level Variables

- *Student's gender (GEN)*: gender of student is a main variable of interest. In the models, male was coded as 0 and female was coded as 1.

- *Socio-economic status (SES) of students*: the Home educational resources scale, which is provided by the TIMSS administration, was used as a proxy of individual student's SES.
- *Outcome variables*: The main outcome variable for the study is overall mathematics performance of students as indicated by a set of five plausible values.

School-level Variables

- *Safe and Orderly School (MTBGSOS):* This variable is aggregated from teacher responses in the same schools. Teachers were asked to answer what extent they agree or disagree with the following statements regarding disciplined climate in their schools using a four-scale response (agree a lot, agree a little, disagree a little, and disagree a lot):

    a) This school is located in a safe neighborhood

    b) I feel safe at this school

    c) This school's security policies and practices are sufficient

    d) The students behave in an orderly manner

    e) The students are respectful of the teachers

    f) This school has clear rules about student conduct

    g) This school's rules are enforced in a fair and consistent manner.

  This variable was scored using the IRT partial credit model developed by the TIMSS administration and standardized so that the mean is 10 and the standard deviation is 2 across all TIMSS participating countries. Higher scores mean very safe and orderly schools.

- *Socio-economic Status (SES) of schools (MEAN SES)*: This variable is aggregated from the students' home educational resources scale (student SES) within each school to represent Socio-economic Status (SES).
- *Communication among Teachers (TCOM)*: This is a derived variable from teacher responses that was created for the purpose of the study. I selected items that capture *communication among teachers* and conducted the exploratory IRT[15]. Items with high factor loadings (higher than .6) were selected to estimate *communication among teachers* and were scored using an IRT GRM. Teachers were asked how often they have the following types of interaction with other teachers for the below items by selecting from four answer categories (very often, often, sometimes, and never or almost never). The sub-items were:
    - f) Discuss how to teach a particular topic
    - g) Collaborate in planning and preparing instructional materials
    - h) Share what I have learned about my teaching experiences
    - i) Work together to try out new ideas
    - j) Work as a group on implementing the curriculum

  This variable score was standardized so that the mean is 10 and the standard deviation is 2 for all sampled teachers in the United States. Higher scores indicate that the teacher interacts more frequently with other teachers. The estimated reliability (Cronbach's alpha) for the *communication among teachers* variable was .91.

[15] The R output for the exploratory IRT is in Appendix A.

- *Teacher support of student participation (TSUP):* This is a derived variable from teacher responses that was created for the purpose of this study. Though TIMSS administered items concerning teacher instructional practices, TIMSS did not provide any latent variables for those items. Thus, I selected items that capture *teacher support of student participation* and conducted the exploratory IRT[16]. Items with high factor loadings (higher than .6) were selected to estimate *teacher support of student participation* and were scored using an IRT GRM. Teachers were asked how often they did the following while teaching the class by selecting from four answer categories (very often, often, sometimes, and never or almost never):

  a) Ask students to explain their answers

  b) Encourage classroom discussions among students

  c) Ask students to decide their own problem solving procedures

  d) Encourage students to express their ideas in class

  This variable was standardized so that the mean is 10 and the standard deviation is 2 for all sampled teachers in the United States. Higher scores indicate more frequent teacher support for student participation. The estimated reliability (Cronbach's alpha) for the *teacher support of student participation* variable was .77.

- *Shortage of teachers (SHORTT):* This is a derived binary variable from school principal responses that was created for the purpose of the study. School principals were asked how difficult it has been this school year to fill 8th grade

[16] The R output for the exploratory IRT is in Appendix A.

mathematics and science teaching vacancies. If they said that a) there were no vacancies for mathematics and science subjects or b) it is easy to fill vacancies for mathematics or sciences, their responses were coded as 0 (no shortage of teachers). If a school principal answered that a) it is somewhat difficult to fill or b) it is very difficult to fill, they were coded as 1 (schools with a shortage of teachers).

Model Specification

At the student-level (level I), individual students' mathematics achievement scores were modeled as a linear relationship of students' gender and SES.

Level I: Student-level model

$$Y_{ij} = \beta_{0j} + \beta_{1j} Gender_{ij} + \beta_{2j} SES_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

The level I covariates ($SES_{ij}$and $Gender_{ij}$) are expressed as deviations from the school means using group-mean centering in the level I model.

School-level models are as follow.

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\boldsymbol{MTBGSOS}_{1j} + \gamma_{02}\boldsymbol{MEAN\ SES}_{2j} + \gamma_{03}\boldsymbol{TCOM}_{3j} + \gamma_{04}\boldsymbol{TSUP}_{4j} + \gamma_{05}\boldsymbol{SHORTT}_{5j} + u_{0j},$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\boldsymbol{MTBGSOS}_{1j} + \gamma_{12}\boldsymbol{MEAN\ SES}_{2j} + \gamma_{13}\boldsymbol{TCOM}_{3j} + \gamma_{14}\boldsymbol{TSUP}_{4j} + \gamma_{15}\boldsymbol{SHORTT}_{5j} + u_{1j},$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}\boldsymbol{MTBGSOS}_{1j} + \gamma_{22}\boldsymbol{MEAN\ SES}_{2j} + \gamma_{23}\boldsymbol{TCOM}_{3j} + \gamma_{24}\boldsymbol{TSUP}_{4j} + \gamma_{25}\boldsymbol{SHORTT}_{5j} + u_{1j},$$

$$\text{Var}\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{bmatrix} = \begin{bmatrix} \tau_{00} & \tau_{01} & \tau_{02} \\ \tau_{10} & \tau_{11} & \tau_{12} \\ \tau_{20} & \tau_{21} & \tau_{22} \end{bmatrix}$$

The matrix on the right-hand side is a variance-covariance matrix. The diagonal terms are the variances of the random effects, where τ_{00} represents the variance in school-mean achievement scores, τ_{11} represents the variance in gender contrasts across schools, and τ_{22} represents the variance in SES-Achievement slopes across schools.

The off-diagonal terms are covariances, where τ_{01} (and τ_{10}) capture the covariance between u_{0j} and u_{1j}, i.e., Cov(u_{0j}, u_{1j}); and τ_{02} (and τ_{20}) capture the covariance between u_{0j} and u_{2j}, i.e., Cov(u_{0j}, u_{2j}).

Among the parameters in the above matrix, the parameters of primary interest with respect to my analyses are the variance terms along the diagonal, particularly, τ_{11}, i.e., the variance in gender contrasts.

When predictors are entered into the level-2 equations, the diagonal terms represent the remaining variance in school-mean achievement, gender contrasts and SES-Achievement slopes, and the off-diagonal terms represent covariances conditional on the predictors in the level-2 equations.

In the three equations above, school-level variables were included to explain school variability in gender gaps. Predictors in the level II model are expressed as deviations from the grand mean (grand-mean centering). The coefficients γ_{1m} (*m*=1, …, 5) represent the relationships between the within-school gender gap and school-level factors.

Results

<u>*Descriptive Statistics for the Variables*</u>

Descriptive statistics for the variables included in the model are summarized in Table 5.6.

Table 5.6. Student and School Variables (TIMSS 2015 Advanced)

Variable	M	SD	Min	Max
Safe and Orderly School	10.38	2.16	4.03	12.99
Socio-economic Status (SES) of schools	10.23	1.23	6.48	12.63
Communication among Teachers	9.55	1.82	5.45	15.54
Teacher support of student participation	10.37	1.82	5.19	12.44
Shortage of teachers	0.76	0.42	0	1

The *safe and orderly school* variable was scaled such that its mean is 10 and standard deviation is 2 across all TIMSS participating countries. The US school sample used in the study has a weighted mean of 10.38 and a standard deviation of 2.16 for the variable. The Socio-economic Status (SES) of schools variable was aggregated from the student SES variable and the mean of US school sample has a weighted mean of 10.23 and a standard deviation of 1.23. The *communication among teachers* and *teacher support of student participation* variables have weighted means close to 10 and standard deviations close to 2. *Shortage of teachers* has a weighted mean of 0.76 and a standard deviation of 0.42 for the all sampled US schools.

<u>*Correlation among the Variables*</u>

A bivariate correlation analysis with the two variables included in the level 1 model – student gender and student SES- indicates that those two variables are not correlated (r = .02, p >.38).

Unweighted bivariate relationships between the level 2 variables were estimated (Table 5.13). Table 5.13 is the correlation matrix for these variables. The correlation coefficients of the level-2 variables ranged from -.11 (between *school SES* and *shortage of teachers*) to .25 (between *safe and orderly school* and *school SES*). *Safe and orderly school* is positively correlated with *school SES*, *communication among teachers* and *teacher support of student participation.* *Communication among teachers* has a positive bivariate correlation with *teacher support of student participation* (r = .19).

Table 5.7. Correlation among Variables (TIMSS 2015 Advanced)

Variable	*Safe and Orderly School*	*Socio-economic Status (SES) of schools*	*Communication among Teachers*	*Teacher support of student participation*	*Shortage of teachers*
Safe and Orderly School	1.00	0.25**	0.12	0.12	-0.03
Socio-economic Status (SES) of schools		1.00	-0.03	0.002	-0.11
Communication among Teachers			1.00	0.19*	-0.07
Teacher support of student participation				1.00	0.1
Shortage of teachers					1.00

Note: †p<.10. *p<.05. **p<.01. ***p<.001.

Results for the Model for Question 2

The results of the Model for question 2 are presented in Table 5.8.

Table 5.8. Results of HLM analysis (TIMSS 2015Advanced)

Variables	Coefficient	*SE*	*t*	*p*
Fixed Effects				
Mathematics achievement				
Intercept	470.06	5.11	91.94	<.001
MTBGSOS	-0.6	2.33	-0.26	0.8
MEAN SES	34.52	4.04	8.55	<.001
TCOM	1.33	2.61	0.51	0.61
TSUP	-2.98	2.89	-1.04	0.3
SHORTT	6.57	11.76	0.56	0.58
Gender				
Intercept	-26.78	4.46	-6	<.001
MTBGSOS	6.47	1.99	3.25	0.001
MEAN SES	-4.55	3.7	-1.23	0.22
TCOM	3.19	2.2	1.44	0.15
TSUP	1.97	3.18	0.62	0.54
SHORTT	18.27	9.89	1.85	0.07
Student SES				
Intercept	4.51	1.14	3.94	<.001
MTBGSOS	-0.32	0.58	-0.55	0.59
MEAN SES	-0.32	0.97	-0.33	0.74
TCOM	1.28	0.6	2.15	0.04
TSUP	-0.37	0.66	-0.56	0.58
SHORTT	0.02	3.12	0.008	0.99

Random Effects[17]	Variance Component	*SD*		*P*
Mathematics achievement	3460.42	58.83	1548.7	<.001
Gender contrast	787.93	28.07	281.94	<.001
Student-level Effect	4986.26	70.61		

Gender contrast variance accounted for by L2 predictors = $\frac{(1144.4-787.93)}{1144.4} = 0.31$

[17] HLM analysis indicates that the random effect for the SES slope is not statistically significant at .05 alpha level and was removed from the analysis.

The variance of gender contrast for this model is 787.93, whereas the gender contrast variance is 1144.4 in the question 1 model. Approximately 31% of gender contrast variance is accounted for by level 2 predictors in the model.

The grand mean of overall mathematics achievement scores for our population of schools is estimated at 470.06 ($p < .001$). Among the variables in the model, only *school SES* was significantly related to the school mean of mathematics achievement scores. For instance, as *school SES* increases by 1 unit, the average overall mathematics achievement score increases by 34.52 points ($p < .001$), holding constant other variables.

For TIMSS 2015 Advanced data, only one variable in the model is related to the gender contrast: *safe and orderly school* (MTBGSOS). In schools where teachers think the school is safe and orderly, the gender gaps favoring male students tend to be lower. C*ommunication among teachers* (TCOM), *teacher support of student participation* (TSUP), and *shortage of teachers* (SHORTT) show positive interaction effects on the gender contrast, although none of these are statistically significant (*ps* > .07).

To facilitate the interpretation, consider two schools whose values for other predictors (MEAN SES, TCOM, TSUP, and SHORTT) are equal to the grand means of those predictors except for the *safe and orderly school* variable. One school (School A) has a value of *safe and orderly school* variable that is one standard deviation (2.16) above the grand mean for *safe and orderly school* variable. The other school (School B) has a value of *safe and orderly school* variable that is one standard deviation (2.16) below the grand mean for *safe and orderly school* variable. For School A, the expected gender gap would be -26.78 + 6.47* (2.16) = -12.8, indicating that male students are expected to outperform female students by 12.8 points in School A. For School B, the expected gender gap would be -26.78 + 6.47* (-2.16) = - 40.76, indicating that male students are expected to outperform female students by 40.76 points in School B. Considering that the student-level random effect is 74.05 in the null model (see Chapter 4), the 27.95 (= -12.8 – (-40.76)) points difference is about 38% of the standard deviation of the level 1 random effect. A 95% confidence interval around the point estimate of *safe and orderly school* is from 2.57 to 10.37, indicating that we are 95% confident that our interval contains or captures the true value of the coefficient for the *safe and orderly school* predictor variable. Note that the range of plausible values for the true value of the coefficient of interest spans only positive values.

For the student SES slope, only *communication among teachers* variable shows a positive interaction effect. As the *communication among teachers* increases by 1 unit, the effect of student SES on the student mathematics achievement also increases by 1.28 (p = .04), holding constant other variables in the model. That is, a student's SES has a stronger impact on mathematic achievement in schools where frequent teacher communication takes place.

Discussion and Implications

While a considerable amount of past research has focused on student and school factors associated with gender differences in mathematics performance, little research has investigated within-school gender gap variability in mathematics achievement. In this chapter, school factors that are associated with within-school gender gap in mathematics were investigated using two different samples of students in the US. These two samples differ in two aspects: school grade and achievement level. Eighth grade students taking the TIMSS 2015 Regular assessment were sampled to represent overall 8th grade students in the US. On the other hand, 12th grade students taking the TIMSS 2015 Advanced assessment were sampled to represent high performing 12th grade US students in mathematics.

<u>Summary of Findings</u>

- 8th grade student
    - *Instruction affected by mathematics resources shortage,* school mean SES, and *communication among teachers* are positively related to school mean mathematics achievement scores.
    - *Shortage of teachers* is negatively related to school mean mathematics achievement scores.
    - *Instruction affected by mathematics resources shortage* and *teacher confidence in teaching mathematics* are negatively related to the magnitude of the gender contrasts.
    - *Communication among teachers* and *teacher support of student participation* are positively related to the magnitude of the gender contrasts.
- 12th grade student
    - School mean SES is positively related to school mean mathematics achievement scores.
    - *Safe and orderly school* variable is positively related to the magnitude of the gender contrasts.

<u>TIMSS 2015 Regular – 8th grade students</u>

For 8th grade students, the findings show that students' mathematic performance is positively related to the *instruction affected by mathematics resources shortage* variable. In other words, students in schools where instruction was less affected by mathematics resources shortage show better performance in mathematics on average than students in schools with mathematics resources shortage. In addition, male students outperform female students by an even greater degree in schools with mathematics resource shortages. Past research on school resources and student performance has reported that there is no systematic relation between school resources and student performance (Hanushek, 1997). In a review article, Hanushek (1997) analyzed articles about school resources and student performance, where school resources mainly included the real resources of the classroom (teacher-pupil ratio, teacher education, teacher s' years of experience), financial aggregates of resources (expenditure per pupil and teacher salary), and measures of other resources in schools (specific teacher characteristics, administrative inputs, and facilities). In this study, mathematics resources include teacher education and mathematic-related facilities, such as computer software and library resources, but do not include general aspects of school resources, such as teacher-pupil ratio, teachers' years of experience or expenditure per pupil. Thus, the inconsistent findings between past research and this study might be due to the different definitions of school resources. One possible reason why gender gap is related to this variable is that schools with a mathematics resource shortage might be less attentive to gender differences in mathematics performance and consequently allocate fewer resources to closing the gap. However, future study is needed to explain the underlying mechanism of this unexpected finding.

The findings show that for 8th grade students, *teacher confidence in teaching mathematics* is not significantly related to overall students' mathematics performance. However, outperformance of male students in mathematics is magnified in schools where *teacher confidence in teaching mathematics* is higher. This finding was unexpected, as past work as well as intuition suggest that teachers who are more confident in teaching mathematics are more effective by implementing new practices (Gabriele & Joram, 2007; Guskey, 1988) and thus would have higher performing students (Goddard, Hoy & Hoy, 2000). Even though the items to measure teacher confidence in teaching mathematics include teacher practices, self-reported confidence in doing those activities might not capture teachers' actual behaviors. Future study is needed to explain the inconsistent findings between previous studies and the current study.

The *shortage of teachers* variable was not statistically related to gender gap variability across schools in the current study. However, a previous study using PISA (Programme for International Student Assessment) data found that a shortage of teachers was related to the school variability in gender gap in the US (Ma, 2008). Again, these inconsistent results might be due to differences in the definitions of the *shortage of teachers* variable between the two studies. In the current study, the *shortage of teachers* variable was derived from school principal assessments on the difficulties of filling mathematics and science teacher vacancies. Whereas in the previous study, the *shortage of teachers* variable was derived from questions to principals around the level of hindrance to student learning due to a lack of teachers in general. The principals who answered the PISA questionnaire might have focused on the hindrance of student learning due to a lack of teachers, whereas the principals who answered the TIMSS questionnaire might have focused on the difficulty within the administration to fill teacher vacancies without considering whether student learning was hindered. At the very least, the current study suggests that administrative difficulties related to filling teacher vacancies do not contribute to within school gender gap variability in mathematics performance.

One interesting finding of the current study is that *communication among teachers* is related to students' mathematics achievement and a decrease in gender gap in mathematics achievement favoring male students. Indeed, recent research on teacher collaboration and student achievement reports that higher level of teacher collaboration is associated with better student performance (Goddard, Goddard, & Tschannen-Moran, 2007; Ronfeldt, Farmer, McQueen, & Grissom, 2015). In particular, a study of more than 300 public schools (Ronfeldt et al., 2015) found that quality of teacher collaboration was positively related to student achievement in mathematics and reading. The current study contributes to these past findings by showing that teacher communication is positively related to students' mathematics achievement and to decreasing the size of the school gender gap in mathematics achievement favoring male students. One note of caution toward this interpretation, however, is that the results of the analyses do not indicate the direction of the association. The analyses are correlational, so the results should not be interpreted as causation. Future work should seek to examine the direction of the relationship and its underlying mechanisms.

Another interesting factor related to a decreased gender gap in mathematics was *teacher support of student participation*. The school-level HLM results indicate that schools with teachers who are more supportive of student participation tend to have a decreased gender gap in mathematics. In particular, greater teacher support of student participation was associated with females performing better in mathematics. In the current study, teacher support of student participation was operationalized as: asking students to explain their answers, encouraging classroom discussion among students, having students decide their own problem solving procedures, and encouraging students to express their ideas in class. By doing these things, teachers provide opportunities to cognitively elaborate on their own ideas, which promotes students' learning (Baines, Blatchford, Kutnick, 2008). Indeed, a body of past research has reported a positive relationship between students explaining their own ideas and higher academic achievement (e.g., Webb et al., 2019). For example, Ing and colleagues (2015) found that teacher support of student participation has an indirect relationship with student achievement, mediated by student participation. Similarly, the current study found that teacher support of student participation is not directly related to overall mathematic achievement of students. Instead, teacher support of student participation is related to a decreased gender gap favoring male students. It is unclear exactly why this relationship exists. One possibility is that female students benefit more than males from warm, supportive teachers who foster active participation in classrooms. However, as results from the HLM analyses do not indicate causation, future work is needed to investigate the associated mechanism through which the gender gap is decreased by teacher support of student participation.

TIMSS 2015 Advanced – high performing 12th grade students

In contrast to the results from the 8th grade students, findings from the high achieving 12th grade students indicates that only one school-level variable in the model is related to the within-school gender gap in mathematics. Schools in which teachers thought their schools are safe and orderly tend to have a decreased gender gap in mathematics achievement favoring male students. This is consistent with prior work that found that school climate is related to student self-concept (Grobel & Schwarzer, 1982; Hoge, Ismit, & Hanson, 1990) and student achievement (Johnson & Stevens, 2006; Kraft & Papay, 2014). In particular, Johnson and Stevens (2006) investigated the relationship between elementary school teachers' perceptions of their school climate and student achievement. They found that the school mean of teachers' perceptions of school climate is positively related to student achievement, indicating that students perform better in schools where teachers have a positive perception of school climate. The current study found that teachers' perceptions of school climate are not significantly related to overall mathematics achievement of 12th grade high performing students. However, the findings show that teachers' perceptions of school climate are associated with a decrease of gender gap in mathematics performance favoring male students. In schools where teachers perceive school climate as safe and orderly, the gender gap in mathematics achievement favoring male students is smaller than their counterparts.

Unlike the results from TIMSS 2015 Regular, while within-school gender gap in high achieving 12th grade students is not related to teacher practices and characteristics, the within-school gender gap is associated with school climate. The differences between the results from two datasets might be explained in several ways.

First, it is possible that high achieving students taking advanced mathematics classes are already highly motivated in learning mathematics and thus specific teacher practices, such as teacher support of student participation, might not affect individual students' mathematics performance compared to regular students.

Second, TIMSS 2015 Advanced assessed high performing students who are at the end of secondary school year. The achievement of those students is accumulated throughout their schooling. Thus, teachers' instruction at the end of secondary school year might not be directly associated the students' achievement.

Third, it is possible that individual psychological characteristic such as persistence or mathematic anxiety might affect students' mathematics performance of high performing students to a greater degree than specific teacher instructional practices. Previous study by Cambell and Beaudry (1998) used path analysis to investigate high performing students' mathematics achievement and found that persistence and self-imposed pressure are related to mathematics performance of female high achieving students.

Fourth, school climate factors such as gender stereotype might have a stronger relationship to mathematics achievement of high performing female students than instructional practices. In fact, a previous study by Haynes, Mullins, and Stein (2004) found that high performing female students reported higher mathematics anxiety and underestimated their actual ability when exposed to teachers who held the stereotype that females are not good at mathematics. Those female students might internalize gender stereotypes, which can be a psychological blocker to mathematical performance.

The TIMSS 2015 datasets do not provide variables to investigate the above-mentioned reasons of gender gap among high achieving students. Future study needs to explore whether and the extent to which those factors are related to gender gap in high performing students.

In conclusion, I investigated factors that are related to within school gender gap variability using both TIMSS 2015 Regular and TIMSS 2015 Advanced datasets in this chapter. Since teacher-level variables such as instructional practices and teacher collaboration are related to 8^{th} grade students' mathematics gender gap but not to 12^{th} grade high performing students, the next chapter will include teacher level HLM analysis to closely investigate teacher-level variables, focusing on only TIMSS 2015 Regular data.

Chapter 6. Research Question 2: Results of Teacher-level HLM

In Chapter 5, factors associated with the within-school gender gap were investigated using an HLM analysis at the school level. The findings indicate that factors related to the within-school gender gap among 8th grade US schools (TIMSS 2015 Regular) are teacher practice and/or teacher characteristics, such as *communication among teachers, teacher confidence in teaching mathematics,* and *teacher support of student participation*. In the school-level HLM analyses, teachers' responses to the questionnaire were aggregated at the school-level to create school-level variables. However, since most schools in the current data set (US schools who administered the TIMSS Regular in 2015) involve only one or two mathematics teachers, it is possible that the aggregated teachers' answers are not representative of their respective schools (which likely have several mathematics teachers). Thus, a sensible alternative is to use teacher data in the teacher-level analysis rather than to aggregate teacher responses to the school-level. In this chapter, a teacher-level HLM analysis was conducted instead of a school-level HLM to more closely investigate the gender gap within schools.

The TIMSS assessment's data is structured such that students are nested within classrooms, within teachers, and within schools. However, classrooms and teachers do not always correspond one to one, as some teachers teach multiple classrooms. Thus, in order to avoid duplicating responses for teachers who teach more than one class, the level 2 unit is teachers instead of classrooms. Furthermore, given that variables included in the model are based on teacher responses (e.g., communication among teachers), it is appropriate to use a teacher-level model and not a classroom-level model. In this chapter, the fact that teachers are nested within schools is ignored because only one or two teachers are in the same school.

Data

Fourteen teachers from 5 schools were excluded from the analysis, as those schools have a complex class and teacher relationship. For instance, multiple teachers teach the same class in some of these schools. Furthermore, an additional 14 teachers with missing values in teacher-level variables were omitted from the analyses. The total number of teachers and students used for the teacher-level HLM analysis is presented below.

Table 6.1. TIMSS 2015 Regular Sample Sizes (Teacher-level HLM)

	TIMSS 2015 Regular
Number of students	7,816
Number of teachers	312
Number of girls	3,957
Number of boys	3,859

In the previous chapter, both student weight and school weight variables were used in the school-level HLM analysis, where student weight includes the probability of selection of students and classrooms and school weight includes the probability of selection of schools. However, because the teacher-level HLM analysis does not involve the selection of schools and classrooms, the teacher-level HLM analyses in this chapter did not apply any weight variables in the model.

Variables of Interest

Table 6.2 provides a brief description of the student- and teacher-level variables used in the models.

Table 6.2. Student and Teacher Variables (TIMSS 2015 Regular)

Variable Name	Description
Outcome	
Math PVs	Mathematics plausible values for the TIMSS 2015 8th grade students
Student-level	
GEN	Gender of student
SES	Socio-economic status of student
Teacher-level	
TCONF	Teacher confidence in teaching mathematics
TCOM	Communication among teachers
TSUP	Teacher support of student participation
TGEN	Gender of teachers
MAJOR	Major of teachers during post-secondary education
UNDER30	Teacher age under 30
AGE30S	Teacher age between 30 and 39
AGE40S	Teacher age between 40 and 49
OVER50	Teacher age over 50 (reference group)
MEAN SES	Socio-economic Status (SES) of students of teachers

Student Level Variables

The same student variables used in the school-level HLM model were used in the teacher-level HLM model.

- *Gender of student (GEN)*: Gender of student is a main variable of interest. In the model, male students were coded as 0 and female students were coded as 1.
- *Socio-economic status (SES) of students*: The Home Educational Resources scale, provided by the TIMSS administration, was used as a proxy of individual student SES.
- *Outcome variables*: The main outcome variable for the study is overall mathematics performance, as captured by a set of five plausible values.

Teacher-Level Variables

Three teacher-level variables that were used in the school-level analyses (Chapter 5) were re-used in the teacher-level HLM analysis (the same teacher-level variables that were used in Chapter 5).

- *Teacher Confidence in Teaching Mathematics (TCONF)*
- *Communication among Teachers (TCOM)*
- *Teacher support of student participation (TSUP)*

Several additional teacher-level variables were created for the teacher-level HLM analysis. Three variables are based off of a single question asking about the teacher's age.

- *Gender of teacher (TGEN):* In the model, male teachers were coded as 0 and female teachers were coded as 1.
- *Major of teacher (MAJOR):* Teachers were asked what their majors or main area of study was during their post-secondary education. Teachers whose major was mathematics or mathematics education were coded as 1 while all others were coded as 0.
- *Age of teacher – under 30 (UNDER30)*: This variable is dummy coded. Teachers whose age was less than 30 were coded as 1 and others were coded as 0. Teachers whose age was over 50 were used as the reference group.
- *Age of teacher – from 30 to 39 (AGE30S)*: This variable is dummy coded. Teachers whose age was between 30 and 39 were coded as 1 and others were coded as 0. Teachers whose age was over 50 were used as the reference group.

- Age of teacher – from 40 to 49 (*AGE40S*): This variable is dummy coded. Teachers whose age was between 40 and 49 were coded as 1 and others were coded as 0. In the model, teachers whose age was over 50 were used as reference group.
- Teacher-level mean SES *(MEAN SES):* This variable is aggregated from students' home educational resources scale (student SES) to represent teacher-level mean SES of students. This measure is used in the teacher-level HLM model to control for the effect of teacher-level mean SES on the gender gap in mathematics achievement.

Model Specification

At the student level (level I), individual students' mathematics achievement scores are modeled as a linear relationship of students' gender and students' SES.

Level I: Student-level model

$$Y_{ij} = \beta_{0j} + \beta_{1j}Gender_{ij} + \beta_{2j}SES_{ij} + \varepsilon_{ij}, \qquad \varepsilon_{ij}N(0, \sigma^2)$$

β_{0j} indicates the mean mathematics score for 8th grade students in school j. The level I covariates ($SES_{ij}$and $Gender_{ij}$) are expressed as deviations from the school mean using group-mean centering in the level I model.

Teacher-level variables were included in the level 2 (teacher-level) model as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}TCONF_{1j} + \gamma_{02}TCOM_{2j} + \gamma_{03}TSUP_{3j} + \gamma_{04}TGEN_{4j} + \gamma_{05}MAJOR_{5j} +$$
$$\gamma_{06}UNDER30_{6j} + \gamma_{07}AGE30S_{7j} + \gamma_{08}AGE40S_{8j} + \gamma_{09}MEANSES_{9j} + u_{0j},$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}TCONF_{1j} + \gamma_{12}TCOM_{2j} + \gamma_{13}TSUP_{3j} + \gamma_{14}TGEN_{4j} + \gamma_{15}MAJOR_{5j} + \gamma_{16}UNDER30_{6j} + \gamma_{17}AGE30S_{7j} + \gamma_{18}AGE40S_{8j} + \gamma_{19}MEANSES_{9j} + u_{1j},$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}TCONF_{1j} + \gamma_{22}TCOM_{2j} + \gamma_{23}TSUP_{3j} + \gamma_{24}TGEN_{4j} + \gamma_{25}MAJOR_{5j} + \gamma_{26}UNDER30_{6j} + \gamma_{27}AGE30S_{7j} + \gamma_{28}AGE40S_{8j} + \gamma_{29}MEANSES_{9j} + u_{2j},$$

$$\mathrm{Var}\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{bmatrix} = \begin{bmatrix} \tau_{00} & \tau_{01} & \tau_{02} \\ \tau_{10} & \tau_{11} & \tau_{12} \\ \tau_{20} & \tau_{21} & \tau_{22} \end{bmatrix}$$

The matrix on the right-hand side is a variance-covariance matrix. The diagonal terms are the variances of the random effects, where τ_{00} represents the variance in school-mean achievement scores, τ_{11} represents the variance in gender contrasts across schools, and τ_{22} represents the variance in SES-Achievement slopes across schools.

The off-diagonal terms are covariances, where τ_{01} (and τ_{10}) capture the covariance between u_{0j} and u_{1j}, i.e., Cov(u_{0j}, u_{1j}); and τ_{02} (and τ_{20}) capture the covariance between u_{0j} and u_{2j}, i.e., Cov(u_{0j}, u_{2j}).

Among the parameters in the above matrix, the parameters of primary interest with respect to my analyses are the variance terms along the diagonal, particularly, τ_{11}, i.e., the variance in gender contrasts.

When predictors are entered into the level-2 equations, the diagonal terms represent the remaining variance in school-mean achievement, gender contrasts and SES-Achievement slopes, and the off-diagonal terms represent covariances conditional on the predictors in the level-2 equations.

In the three equations above, teacher-level variables were included to try to explain variability in the mathematics achievement gender gap. Predictors in the level II model are expressed as deviations from the grand mean (grand-mean centering). The coefficients γ_{1m} (m=1,…, 9) represent the relationships between the gender gap and teacher-level factors.

Results

Descriptive Statistics for the Variables

Tables 6.3 and 6.4 shows the minimum value, maximum value, mean, and standard deviation for each variable included in the models.

Table 6.3. Descriptive Statistics for Level 1 Variables (TIMSS 2015 Regular)

Variable	M	SD	Min	Max
Student Gender	0.51	0.50	0	1
Socio-economic Status (SES) of student	10.82	1.68	4.23	13.88

Table 6.4. Descriptive Statistics of the Variables (TIMSS 2015 Regular –Teacher-level HLM)

Variable	M	SD	Min	Max
Teacher confidence in teaching mathematics	9.98	1.96	5.7	13.61
Communication among Teachers	10.00	1.98	2.9	12.3
Teacher support of student participation	10.02	2.01	5	12.5
Gender of teacher	0.69	0.46	0	1
Major of teacher	0.68	0.47	0	1
Teacher age - under 30	0.15	0.36	0	1
Teacher age - between 30 and 39	0.30	0.46	0	1
Teacher age - between 40 and 49	0.30	0.46	0	1
Socio-economic Status (SES) of students of teacher	10.81	0.96	8.87	13.25

The *teacher confidence in teaching mathematics*, *communication among teachers*, and *teacher support of student participation variables* were scored such that the means were close to 10 and the standard deviations were close to 2. For this data set, 69% of teachers are female. In terms of teacher age, 15% are under 30, 30% are between 30 and 39, and 30% are between 40 and 49.

Correlation among Variables

Bivariate relationships between level 2 (teacher-level) variables were estimated. Table 6.4 presents the correlation matrix for these variables.

The correlation coefficients of the level-2 variables ranged from -.43 to .4 and are presented in Table 6.5. Some interesting correlations were *teacher confidence in teaching mathematics and communication among teachers* (*r* = .37), indicating that teachers with more confidence in teaching mathematics also communicate with other teachers more. In addition, *teacher confidence in teaching mathematics* was related to *teacher support of student participation* (*r* = .4), indicating that teachers with more confidence in teaching mathematics are more likely to support student participation. Unsurprisingly, teachers whose major was mathematics or mathematics education report higher *teacher confidence in teaching mathematics*, (*r* = .1 between those two variables). *Communication among teachers* is positively correlated with *teacher support of student participation* (*r* = .23), indicating that teachers who reported frequent communication with other teachers are more likely to support student participation. *Teacher support of student participation* has a positive correlation with *gender of teacher* (*r* = .14), indicating that female teachers are more likely to support student participation. It also has a positive correlation with *major of teachers* (*r* = .14), suggesting that teachers with mathematics or education of mathematics tend to support more student participation.

Table 6.5. Correlations among Variables (TIMSS 2015 Regular- Teacher-level HLM)

Variable	*Teacher confidence in teaching mathematics*	*Communication among Teachers*	*Teacher support of student participation*	*Gender of teacher*	*Major of teacher*	*Teacher age - under 30*	*Teacher age - between 30 and 39*	*Teacher age - between 40 and 49*	*Socio-economic Status (SES) of students of teacher*
Teacher confidence in teaching mathematics	1.00	0.37***	0.40***	-0.05	0.1†	-0.02	-0.07	0.06	-0.02
Communication among Teachers		1.00	0.23***	0.05	0.04	0.01	-0.04	0.04	-0.02
Teacher support of student participation			1.00	0.14*	0.14*	0.04	-0.07	0.08	0.03
Gender of teacher				1.00	0.06	0.04	-0.1†	-0.02	0.12*
Major of teacher					1.00	0.06	-0.05	0.06	-0.02
Teacher age - under 30						1.00	-0.28***	-0.28***	0.01
Teacher age - between 30 and 39							1.00	-0.43***	-0.1†
Teacher age - between 40 and 49								1.00	-0.03
Socio-economic Status (SES) of students of teacher									1.00

Note: †p<.10. *p<.05. **p<.01. ***p<.001.

Results of the Null Model

First, the null model was constructed to examine intraclass correlations (ICC). Table 6.6 shows the results of the null model with overall mathematics performance scores as the outcome. The estimate of the grand mean of the mathematics achievement scores for our population is 516.49 (p <.001). Teacher-level variability in mathematics proficiency is estimated at 3547.32 and there is significant variation in mean mathematics proficiency, $\chi^2 = 8225.15$, $p < .001$. Student-level variability is estimated at 3314.88. The ICC coefficient is 0.52, indicating that 52% of the total variance is between-teachers. The results of the null model also show an estimate of the reliability of sample means. The reliability of the estimate is .96, indicating that the sample means are quite reliable as indicators of the true mean. Under the normality assumption, the 95% plausible value range for the mean achievement of teachers' students is 516.49 ± 1.96*(57.57) = (403.65, 629.33).

Table 6.6. Parameter Estimates for the Null Model (TIMSS 2015 Regular- Teacher-level HLM)

	Coefficient	*SE*	*t*	*p*
Fixed Effects				
Intercept (γ_{00})	516.49	3.46	149.07	<.001
Random Effects	Variance Component	*SD*	χ^2	*p*
Teacher-level Effect (τ_{00})	3547.32	59.56	8225.15	<.001
Student-level Effect (σ^2)	3314.88	57.57		

$$ICC = \frac{3547.32}{3547.32+3314.88} = 0.52$$

Results for the Random-coefficients Regression Model

Before conducting teacher-level HLM analyses, a random-coefficients regression model was examined. In a random-coefficient regression model, relationships between the outcome variable and predictors used in the level 1 model vary randomly across the level 2 units where there are no teacher-level variables included. The results of the random-coefficients regression model are presented in Table 6.7.

Table 6.7. Parameter Estimates for the Random-coefficients Regression Model (TIMSS 2015 Regular- Teacher-level HLM)

Variables				
	Coefficient	*SE*	*t*	*p*
Fixed Effects				
Intercept (γ_{00})	516.52	3.46	149.10	<.001
Within-School Effects				
Gender (γ_{10})	-7.88	1.55	-5.08	<.001
SES (γ_{20})	7.36	0.56	13.26	<.001
Random Effects	Variance Component	*SD*	χ^2	*P*
School-level Effect(τ_{00})	3553.31	59.61	8627.83	<.001
Gender contrast (τ_{11})	82.63	9.09	336.00	.16
SES slope (τ_{22})	15.15	3.89	350.79	.06
Student-level Effect (σ^2)	3127.68	55.93		

Student-level variance accounted for by L1 predictors = $\frac{(3314.88-3127.68)}{3314.88}$ = 0.06

Overall, the random-coefficients regression model accounted for 6% of student-level variability, compared to the null model, by including student gender and SES. The grand mean of mathematics achievement scores for our population is estimated at 516.52 (p <.001). Of the 312 teachers, the number of teachers whose male students outperform female students is 190, whereas the number of teachers whose female students outperform male students is 122[18]. As for the slope, as student's SES increases by 1 unit, we expect an increase of 7.36 points in a student's overall mathematics achievement score (p < .001). The mean gender difference in

[18] Distribution of within-school gender gap in mathematics achievement across all teachers is located in Appendix B (Figure B-24).

overall mathematics performance across teachers is estimated at -7.88, indicating that male students significantly outperform female students ($p < .001$). However, the variance component for the gender contrast is 82.63, which is not statistically significant, $\chi^2 = 336.00$, $p = .16$ (see Table 6.6). This indicates that gender differences in mathematics performance do not significantly vary across teachers. Under the normality assumption, the 95% plausible value range for the gender gap slope is -7.88 ± 1.96*(3.89) = (-15.5, -0.26). This range indicates that there is smaller variability in the size of the gender gap among teachers compared to that of schools (see Chapter 5). Since the variances of the slope for gender and student SES were not statistically significant at the 0.05 alpha level (from the random-coefficients regression model), I next conducted a teacher-level HLM analysis using a model with a nonrandomly varying slope (Raudenbush & Bryk, 2002). In this model, both gender and student SES slopes vary across teachers but the variations are not random.

Results for the Teacher-level HLM

The results of the teacher-level HLM analysis are presented in Table 6.8

Table 6.8. Results of teacher-level HLM analysis (TIMSS 2015 Regular)

Variables	Coefficient	*SE*	*t*	*p*
Fixed Effects				
mathematics achievement				
Intercept	516.60	2.36	219.19	<.001
TCONF	4.08	1.4	2.92	0.004
TCOM	-0.24	1.3	-0.19	0.85
TSUP	1.17	1.33	0.88	0.38
TGEN	4.26	5.23	0.82	0.42
MAJOR	2.71	5.14	0.353	0.6
UNDER30	4.92	7.7	0.64	0.52
AGE30S	-3.46	6.5	-0.53	0.59
AGE40S	-6.55	6.51	-1.01	0.32
MEAN SES	45.55	2.52	18.05	<.001
Gender				
Intercept	-7.92	1.51	-5.24	<.001
TCONF	-1.15	0.81	-1.43	0.16
TCOM	-0.15	0.76	-0.2	0.84
TSUP	1.73	0.91	1.89	0.07
TGEN	-7.03	3.26	-2.16	0.03
MAJOR	3.76	3.37	1.12	0.27
UNDER30	-8.95	4.44	-2.02	0.04
AGE30S	-5.09	3.88	-1.31	0.19
AGE40S	-4.07	3.84	-1.06	0.29
MEAN SES	-1.51	1.51	-1	0.32
Student SES				
Intercept	7.57	0.52	14.54	<.001
TCONF	-0.69	0.33	-2.13	0.04
TCOM	0.48	0.28	1.73	0.09
TSUP	0.34	0.29	1.2	0.23
TGEN	0.99	1.04	0.95	0.34
MAJOR	-0.94	1.16	-0.82	0.42
UNDER30	0.42	1.57	0.27	0.79
AGE30S	-0.83	1.34	-0.62	0.54
AGE40S	-0.79	1.26	-0.63	0.53
MEAN SES	1.14	0.62	1.85	0.07

Random Effects[19]	Variance Component	*SD*	χ^2	*P*
Mathematics achievement	1542.51	39.27	3665.13	<.001
Student-level Effect	3173.21	56.33		

The grand mean of overall mathematics achievement scores for our population is estimated at 516.6 (p <.001). *Teacher confidence in teaching mathematics* and teacher-level mean SES were significantly related to the mean of mathematics achievement scores at the 0.05 alpha level. The average mathematics achievement score for students whose teachers are confident in teaching mathematics tends to be higher than that of students with less confident teachers. In particular, as teacher confidence increases by 1 unit, average mathematics achievement scores increases by 4.08 points (p = .004), holding constant other variables in the model. Also, as teacher-level mean SES (MEAN SES) increases by 1 unit, average overall mathematics achievement scores increases by 45.55 points (p <.001), holding constant other variables in the model.

Regarding the gender contrast, three variables in the model seem to be related to the gender gap at the 0.1 significance level. The intercept for the gender contrast is -7.92. This means that the expected gender gap for a teacher whose values for the predictors (TCONF, TCOM, TSUP, TGEN, MAJOR, UNDER30 AGE30S, AGE40S, and MEAN SES) are equal to the grand means of those predictors is -7.92, indicating males outperform females. The gender of teachers (TGEN) and teacher age under 30 (UNDER30) are negatively associated with the gender contrast, while *teacher support of student participation* (TSUP) has a positive association.

[19] HLM analysis indicates that the random effects for the gender contrast and the SES slope are not statistically significant at .05 alpha level, thus they were removed from the model.

(note: negative values for the gender contrast indicate male outperformance, while positive values indicate female outperformance).

Female teachers are related to an increased gender gap favoring male students by 7.03 points (p = .03). Compared to teachers over 50, teachers under age 30 tend to be associated with an increased gender gap favoring male students, by 8.95 points (p = .04). For *teacher support of student participation* (TSUP), a 1 unit increase is associated with a 1.73 point decrease in the gender gap favoring male students, a marginally significant effect (p = .07), holding constant other variables in the model.

Although it is a marginally significant effect at the .05 alpha level, we can use a more liberal alpha level of .1 for *teacher support of student participation.* To illustrate these findings, consider two teachers whose values for other predictors (TCONF, TCOM, TGEN, MAJOR, UNDER30, AGE30S, AGE40S, and MEAN SES) are equal to the grand means of those predictors except for *teacher support of student participation*. One teacher (Teacher A) has a value for *teacher support of student participation* that is one standard deviation (2.01) above the grand mean for that variable. The other teacher (Teacher B) has a *teacher support of student participation* variable that is one standard deviation (2.01) below the grand mean for that variable. For Teacher A, the expected gender gap would be -7.92 + 1.73* (2.01) = -4.44, indicating that male students are expected to outperform female students by 4.44 points. For Teacher B, the expected gender gap would be -7.92 + 1.73* (-2.01) = -11.4, indicating that male students are expected to outperform female students by 11.4 points. Considering that the student level random effect is 57.57 in the null model, the 6.95 (= - 4.47 – (-11.4)) points difference is about 12% of the standard deviation of the level 1 random effect. A 95% confidence interval around the point estimate of *teacher support of student participation* is from -0.05 to 3.51, indicating that we are 95% confident that our interval contains or captures the true value of the coefficient for *teacher support of student participation*. Note that the range of plausible values for the true value of the coefficient of interest includes a negative lower bound (with a small magnitude), though most of the range spans positive values.

This finding shows that higher *teacher support of student participation* is related to a decrease in the gender gap in mathematics achievement favoring male students, when we use a more liberal alpha level at .1 (p =.07).

For the slope of student SES, three variables in the model are related with the gender gap at the 0.1 significance level. *Communication among teachers* (TCOM) and teacher-level mean SES (MEAN SES) are positively associated with the slope of student SES. To illustrate, as *communication among teachers* increases by 1 unit, the effect of student SES on mathematics achievement scores increases by 0.49 (p = .09); as teacher-level mean SES increases by 1 unit, the effect of student SES on mathematics achievement scores increases by 1.14 (p = .07). In contrast, *teacher confidence in teaching mathematics* (TCONF) has a negative association with the slope of student SES such that as *teacher confidence* increases by 1 unit, the effect of student SES on mathematics achievement scores decreases by 0.69 (p = .04).

Additional Analyses

Both the results of the school-level HLM and the teacher-level HLM demonstrate that *teacher support of student participation* is associated with the within school (or teacher) gender gap (note: *teacher support of student participation* is significantly related to the gender gap in the teacher-level HLM analysis with a more liberal alpha level of .1). In both models, teacher support of student participation is positively related to the gender gap, such that greater teacher support of student participation is associated with a decreased gender gap favoring males. To more closely investigate the effect of teacher support of student participation, I selected teachers who teach two mathematics classes to conduct additional analyses. In particular, I investigated whether two classes taught by the same teacher look similar in terms of the gender gap.

In this sample, 62 teachers have two classrooms. In order to calculate *within class gender gap scores*, the five plausible values for overall mathematics performance for each student were averaged to calculate a single value for students' mathematics score. Then, the class means for male and female students' mathematics scores were calculated. Next, the class mean for male students was subtracted from the class mean of female students to calculate a *within class gender gap sco*re (Note: positive values indicate that females outperformed males, whereas negative values indicate male outperformance; see Figure 6.1). Since class size varies across classes and gender gaps calculated from classes with a small number of students might exhibit more extreme values, I investigated the relationship between class size and gender gap. The analysis shows that class size does not appear to be related to the magnitude of the gender gap (r = -.11, p = .25)[20].

In Figure 6.1, each teacher from the sample is represented by a vertical bar. The end points of each bar represent the *within class gender gap scores* of the teacher's two classes. The length of each bar represents the difference in scores between the two classes. The bars are ordered by length. As the figure demonstrates, there is significant variation in the size of the gender gap between classes taught by the same teacher. Thus, it is difficult to conclude that teacher specific variables are associated with the gender gap in a uniform way.

[20] Scatterplot of class size and gender gap is presented in Appendix B (Figure B.23).

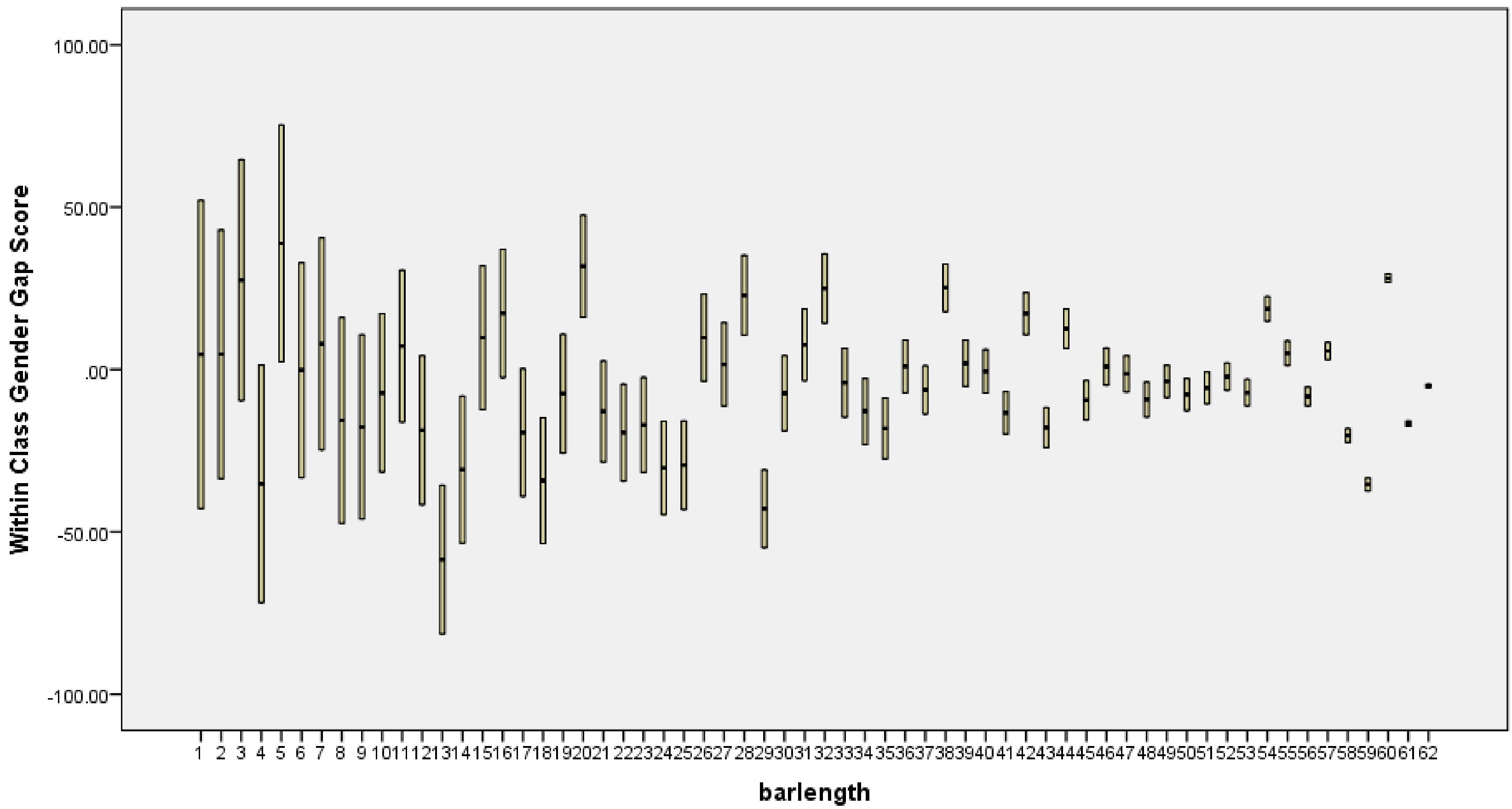


Figure 6.1. Bar Graph of the Within-class Gender Gap of Two Classes Taught by a Same Teacher (TIMSS 2015 Regular)

To further investigate variability of the within class gender gap, teachers were categorized based on the size of the gender difference between their two classes. Teachers whose within class gender gap scores differed by 10 points or more were categorized as "*higher gap diff group* (*N* = 34)," whereas teachers whose within class gender gap scores differ by less than 10 points were categorized as "*lower gap diff group* (*N* = 28)."

Table 6.9. Descriptive Statistics of Teacher Characteristics by High and Low Gap Diff Groups

	Higher Gap Diff Group (*N* = 34)	Lower Gap Diff Group (*N* = 28)
Class mean math scores	511.80 (*SD* =67.09)	524.45 (*SD* =61.95)
Percentage of female teachers	88%	68%
Percentage of teachers with mathematics major	70%	79%

Descriptive statistics demonstrate that teachers in each group differ in several aspects. Teachers in the *higher gap diff group* tend to have lower class mathematics mean scores (M = 511.8, SD = 67.09) compared to their counterparts (M = 524.45, SD = 61.95). Also, teachers in the *higher gap diff group* tend to have a higher percentage of female teachers (88%) compared to the *lower gap diff group* (68%). Finally, 70% of the teachers in the *higher gap diff group* majored in mathematics or mathematics education, whereas 79% of teachers in the *lower gap diff group* majored in math or math education.

Table 6.10. Percentage of English Language Speaking Students by High and Low Gap Diff Groups

English speakers	Higher Gap Diff Group (*N* = 34)	Lower Gap Diff Group (*N* = 28)
More than 90%	68.8%	81.5%
76% to 90%	9.3%	11.1%
Less than 75%	21.9%	7.4%

I next investigated if there were differences in the schools taught by high and low gap teachers. Both groups also differ in terms of several school characteristics. Teachers in the *higher gap diff group* tend to be in schools that have more non-native English speakers, as reported by school principals. For example, teachers in the *low gap diff group* were more likely to be in schools with predominately native English speakers (93% of schools have more than 75% native English speakers), whereas teachers in the *higher gap diff group* were in schools with only 78% of students being native English speakers.

Table 6.11. Location of Schools by High and Low Gap Diff Groups

	Higher Gap Diff Group (*N* = 34)	Lower Gap Diff Group (*N* = 28)
Urban or suburban	37.5%	22.2%
Medium size city	31.3%	22.2%
Small town or remote rural	31.2%	55.6%

In addition, schools with teachers in the *lower gap diff group* tend to be located in small towns or rural areas, whereas schools of teachers in the *higher gap diff group* are equally distributed in large, medium, and small sized towns. This might suggest that classrooms in small towns are more homogeneous, contributing to a smaller gender gap

Discussion and Implications

The results from the school-level HLM in Chapter 5 showed that teacher-level variables such as teacher communication with other teachers and instructional practices are associated with a within-school gender gap in mathematics achievement. In the present chapter, a teacher-level explanatory model was constructed to investigate the extent to which teacher-level variables are associated with gender differences in mathematics achievement. In particular, a teacher-level HLM analysis was conducted on variables such as teachers' gender and age, confidence, and instructional practices.

The results showed that three teacher-level variables are associated with a gender gap in mathematics achievement: gender of teacher, age of teacher, and teacher support of student participation. Specifically, female teachers and teachers under age 30 tend to have classes with an increased gender gap in mathematics performance favoring male students. In contrast, teachers who support student participation tend to have classes with a decreased gender gap in mathematics performance, which is consistent with the results from the school-level HLM.

The present results fit with past research on the influence of teacher gender on students' performance, though admittedly this work is somewhat mixed. For example, some research found that female teachers promoted better reading and language learning performance, although interestingly they were not related to students' mathematics learning (Chudgar & Sankar, 2008; Krieg, 2005; Lam, Tse, Lam, & Loh, 2010). Other studies found that teachers that are the same gender as their students had a positive effect on students' achievement (i.e., female students perform better when they are taught by female teachers; Dee, 2006; Michaelowa, 2001). Recently, Beilock and colleagues (2010) investigated elementary school teachers and their students' mathematics anxiety. Their findings indicate that while mathematics anxiety in female teachers affects female students' mathematics stereotypes and low achievement in mathematics, this relationship does not hold for male students. Although it is an empirical question whether this finding would still hold for *secondary* school teachers, it is possible that these findings bear on the present study. Indeed, the current study found that female teachers are associated with an increased gender gap favoring male students, which may in part be caused by female teachers' own math anxiety negatively influencing their female students. This is a possible interpretation and should be viewed with caution. Future work should seek to address this question directly by examining whether female teachers' mathematics anxiety in secondary schools influence their female students' mathematics anxiety.

In the current study, I found that teacher age was associated with an increased gender gap in mathematics performance. Past research on teacher age indicates that older teachers tend to show more trust in their students, compared to younger teachers (Van Houtte, 2007). Since trust is "assured reliance on character, ability, strength" (defined by Merriam & Webster), teachers' trust in students is a critical component to cultivating effective learning environments. The finding that teachers who are older (age over 50) tend to have a decreased gender gap in mathematics achievement might be related to older teachers having more trust in their students and thus providing a more effective learning environment.

Consistent with the results from the school-level HLM (Chapter 5), teacher support of student participation is associated with the gender gap in the teacher-level HLM analysis. Specifically, greater teacher support of student participation is related to a decreased gender gap favoring male students. By encouraging student participation, teachers provide students with opportunities to explain their own ideas to others, fostering students' learning. A number of past studies show that male students tend to be dominant in class participation (e.g., Swann & Graddol, 1988, Good, 1981; Fassinger, 1995a, 1995b). Moreover, female students may be less likely to speak out in class because they tend to be more self-conscious about making mistakes and embarrassing themselves than male students tend to be. Thus, teachers' support and encouragement might benefit female students more by creating a warm environment where students feel their opinions are welcomed and valued, resulting in an increased propensity to participate in class activities and consequently improve their mathematics performance.

This chapter also includes additional analyses investigating a subsample of teachers who teach two classes in order to probe the complexity of the gender gap in mathematics performance. The results indicate that the within class gender gap scores vary significantly between the two classes taught by a given teacher. The analysis revealed that teachers with large differences between their two classes (i.e., *high diff gap group*) tend to have lower class mean mathematics scores and are more likely to be female. There are a couple of school characteristics - percentage of non-English speaking students and school location - that differ between teachers in the *high diff gap group* (i.e., large difference of within class gender gap score between the teacher's two classes) and *low diff gap group*. The results from this additional analysis demonstrate the complexity of investigating the within school gender gap in mathematics performance. In order to comprehensively understand it, future studies are needed to isolate these novel and potentially associated factors.

Chapter 7. Conclusion

In K-12 education, female students have been widely reported to underperform on standardized mathematics tests relative to male students. This research goes back decades, though recent studies have found that the gender gap in mathematics seems to be decreasing (e.g., Else-Quest, Hyde, & Linn, 2010; Maccoby & Jacklin, 1974). Nevertheless, results from several large-scale standardized mathematics assessments (U.S. National Assessment of Educational Progress, PISA, and TIMSS) indicate that the gender gap in mathematics still exists (McGraw, Lubienski, & Strutchens, 2006; OECD, 2000, 2004; TIMMS, 2000a and b, 2003).

A considerable body of research has investigated individual (e.g., math anxiety and math self-efficacy) and social (e.g., gender stereotype) factors related to the gender gap in mathematics performance. Little research, however, has been directed at school variability in gender differences in mathematics performance and the impact that school environments might have on these differences.

The purpose of the current study was to investigate school variability in gender differences in mathematics achievement and examine associated school-level factors. Investigating within-school gender differences is important because a student's reference group is other students from the same school. In this way, the school may directly influence students' perceptions of mathematics and whether they view mathematics in a gender-stereotyped way. For example, females in a school with a wide gender gap in mathematics performance might have the "males do better at math" stereotype reinforced by school peers, which could hinder them from maximizing their effort in mathematics.

The current study used data from the TIMSS (Trends in International Mathematics and Science Study) 2015 mathematics assessment, an international, large-scale standardized assessment. Since the educational system is inherently hierarchical – students are nested within classrooms and schools – hierarchical linear modeling (HLM) was used as the main analytical method.

Summary of Findings

Question 1. Are there gender differences in mathematics achievement after controlling for students' SES? To what extent do gender differences in mathematics achievement vary across schools?

- Male students perform better than female students in both 8^{th} and 12^{th} grades.
- The magnitude of the gender gap for 12^{th} grade students is greater than that of 8^{th} grade students.
- There is considerable variability between schools in the size of the gender gap in mathematics performance.
- The variability across schools in mathematics performance assessment scores between genders is even greater for 12^{th} grade high achieving students.

Question 2a. What school-level factors are associated with school variability in gender differences in mathematics achievement? (School-level HLM)

- 8^{th} grade students
    - *Instruction affected by mathematics resources shortage*, school mean SES, and *communication among teachers* are positively related to school mean mathematics achievement scores.
    - *Shortage of teachers* is negatively related to school mean mathematics achievement scores.
    - *Instruction affected by mathematics resources shortage* and *teacher confidence in teaching mathematics* are negatively related to the magnitude of the gender contrast.
    - *Communication among teachers* and *teacher support of student participation* are positively related to the magnitude of the gender contrast.
- 12^{th} grade students
    - School mean SES is positively related to school mean mathematics achievement scores.

- *Safe and orderly school* variable is positively related to the magnitude of the gender contrast.

Question 2b. What school-level factors are associated with school variability in gender differences in mathematics achievement? (Teacher-level HLM)

- Gender of teacher, age of teacher, and *teacher support of student participation* were associated with a within-school gender gap in mathematic achievement.
- Female teachers and teachers under age 30 tend to have classes with an increased gender gap in mathematics performance favoring male students.
- Teachers who support student participation tend to have classes with a decreased gender gap in mathematics performance.
- Additional analyses investigating a subsample of teachers who teach two classes indicate that the within class gender gap scores vary significantly between the two classes taught by a given teacher.

The first research question investigated whether there are gender differences in mathematics achievement and to what extent these gender differences vary across schools. In short, the current study found that for students in both 8th and 12th grades, male students perform better than female students. The magnitude of the gender difference for the high performing 12th grade students was greater than that of 8th grade students, which is consistent with prior research (e.g., Hyde, Fennema, & Lamon, 1990).

The current study also found substantial variability between schools in the size of the gender gap in mathematics performance. For 8th grade students, the range of the within-school gender gap was about half of the standard deviation of mathematics performance scores. Variability in the size of the gender gap across schools was even greater for 12th grade high achieving students, as the range was more than one standard deviation.

Given the variability in the size of the within school gender gap in mathematics performance, it was important to investigate related factors using school-level HLM analyses. Findings from the current study indicated that there were several predictors associated with the within-school gender gap in 8^{th} grade, but only one for high performing12^{th} graders. Eighth grade students attending schools that are less affected by a mathematics resources shortage performed better in mathematics than schools with a mathematics resources shortage. In addition, male 8^{th} graders outperformed female 8^{th} graders by an even greater degree in schools with a resource shortage. Furthermore, the gender gap favoring male students was magnified in schools where *teacher confidence in teaching mathematics* was higher. It is unclear why these two predictors – *instruction affected by a mathematics resource shortage* and *teacher confidence in teaching mathematics* - are related to male student outperformance in mathematics assessments. One possible reason is that schools with a mathematics resource shortage might be less attentive to gender differences in mathematics performance and consequently allocate fewer resources to decreasing the differences. It should be noted that teacher confidence was a self-reported construct and not empirically measured. Thus, it is possible that self-reported confidence of teachers might not capture teachers' actual behaviors. Again, future study is needed to explain the underlying mechanism of this unexpected finding.

Two additional factors were found to be related to the within-school gender gap in mathematics: *communication among teachers* and *teacher support of student participation*. Communication among teachers was positively related to students' overall mathematics achievement and a decrease in the size of the gender gap. Teacher support of student participation was also related to a decreased gender gap in mathematics favoring male students. Indeed, greater teacher support of student participation was associated with females doing better in mathematics in most schools. As for the underlying mechanism, it is plausible that schools with teachers who frequently communicate engender a supportive teacher network, which may help keep struggling students from “falling through the cracks” or from having their poor performance ignored. In addition, it is possible that teachers’ support and encouragement of class participation might benefit female students more by creating a warm environment where female students feel their opinions are welcomed and valued, resulting in improved mathematics performance.

In contrast, only school climate was related to the within-school gender gap for 12th grade high achieving students. Schools that teachers felt were safe and orderly tended to have a decreased gender gap in mathematics achievement. Unlike results from TIMSS 2015 Regular, the gender gap was not related to teacher practices and characteristics. It is possible that the current study failed to find a relation because high achieving students taking advanced mathematics classes are already highly motivated. Thus, specific teacher practices, such as teacher support of student participation, might not affect these students as much as regular students. Future work should explore factors associated with the gender gap in high performing students and why high performing students may be different from regular students when it comes to the gender gap in mathematics performance.

In the current study, I found that most variables associated with the within-school gender gap were constructed from the teacher answers in the questionnaire. Since most schools in the sample in TIMSS 2015 Regular data only had one or two mathematics teachers’ responses, it is possible that the teachers’ answers are not representative of the schools, which have dozens of teachers. Thus, it might be more precise to use teacher-level data in a separate, teacher-level analysis than to aggregate teacher responses to the school level. Moreover, research has shown that teaching is the most important in-school factor impacting student learning (Rivkin, Hanushek, & Kain, 2000; Rowan, Correnti & Miller, 2002). As such, teacher-level variables were investigated more closely using a teacher-level HLM analysis. Several teacher-level variables were found to be associated with the within-school gender gap in mathematics achievement. For example, female teachers and teachers under age 30 were more likely to have classes with a wider gender gap. In addition, consistent with results from the school-level HLM, teachers who support student participation were more likely to have classes with a smaller gender gap in mathematics performance. In addition, I also investigated a sub-sample of teachers who have two classes. The analysis showed that there was significant variation in the size of the gender gap between the teachers' two classes (i.e., some teachers have classes with similar size gender gaps while other teachers have one large gap and one small gap class). This illustrates the complexity of the within-school gender gap and suggests the need for future studies.

Limitations

Although the TIMSS 2015 dataset offers some advantages for the purposes of this study, these types of large-scale assessment datasets have inherent limitations.

First, missing responses in the dataset limited the ability to utilize the full sample of data. The current study examined not only student level variables, but also teacher and school level variables. Ideally, the dataset would have provided complete, non-missing responses for all levels. However, due to missing responses in teacher and school principal questionnaires, approximately 20% of data was excluded from the study, reducing sample size and limiting statistical power in data analysis. Future studies should seek to minimize the impact of missing responses so that the full data set can be leveraged to more accurately measure construct and relationships among variables.

Second, the background data for the TIMSS assessment (student, school principal, and teacher questionnaires) are generally based on ”snapshot”, self-report responses. Relying on a single measure for a construct inherently assumes stability of the variables, which may be problematic as some constructs change over time. For example, *communication among teachers* was constructed based on a single question posed to teachers at the end of the school year. It is possible that as the school year progressed, certain teacher became more communicative with their fellow teachers. In addition, self-reported responses can be influenced by social desirability. For example, a teacher might claim that he or she supports student participation more frequently than reality in order to be seen more favorably by others.

A third limitation of the TIMSS assessment is that it only provides a current measure of students' achievement, with no consideration of the students' prior scores. Since student achievement reflects multiple and cumulative environmental effects, it would be more accurate to take into account prior achievement as well.

A fourth limitation of the TIMSS assessment is that the analysis is limited to the questions that were asked in the assessment. I would like to have investigated variables that were not asked in the TIMSS assessment, such as the gender stereotyped attitudes of teachers and parents. In addition, the international dataset does not reflect nation-specific characteristics of the education system. For instance, it would be informative to know the specific level of US mathematics classes (such as pre-algebra or algebra classes) that students were enrolled in. Thus, some variables that might further illuminate the present findings are missing in this large-scale international assessment dataset.

A final limitation of the TIMSS data is that it is cross-sectional and by nature can only provide correlational, not causal, relationships between variables. Being able to make causal inferences and associated underlying mechanism would be a critical benefit to the present analysis. For instance, the current study found that *teacher support of student participation* is related to a decreased gender gap in mathematics performance. It is unclear, however, as to whether teacher support of student participation decreases the gender gap, or the other way around, or if some unmeasured variable is also associated. Teachers may support students' participation more often in classes where female students perform as well as males in mathematics because all students are able to participate in class activities. Similarly, it is possible that high parental expectations and involvement might lead to more teacher support of class participation and promote their students' mathematics achievement.

Future Directions

As indicated in the limitations section, further research is needed to elucidate the underlying causal mechanisms of gender differential performance in mathematics. Three avenues are suggested for future study.

First, in-depth qualitative research would help to more accurately measure constructs of interest. For example, interviews of teachers or a teacher focus group would be a valuable supplement to the teacher questionnaire. It would be particularly useful to know if teachers recognize a gender gap in their classes and what types of effort they make, if any, to rectify it. Classroom observation would be another interesting direction for future study. For one, observation could provide estimates of class behavior that are more objectively accurate than survey responses, which might be influenced by social desirability. For example, we could count the number of times teachers call on students to answer questions as a measure of *teacher support of student participation,* and which students (male or female) they call on. Observational studies can also provide information on classroom dynamics between teachers and students. For example, teachers may be more likely to encourage female students to participate during a small group discussion than during a classroom lecture.

A second promising avenue for future work is longitudinal research to draw causal inferences between variables. For instance, assessments could be administered at the beginning and end of school year. Even though multiple factors influence gender differences in mathematics performance, being able to look at the difference in the size of the gender gap between the two time points would provide further insight into the effect of the teacher and school. Thus, longitudinal studies would help to explore and potentially strengthen claims of a causal relationship between the gender gap in mathematics achievement and teacher-level variables.

A final direction for future study would be to expand to a quasi-experimental design. For example, teacher 'interventions' could be conducted in order to encourage some teachers to foster more participation. To illustrate, one class would have a teacher who is being trained to consciously and frequently encourage student participation, while a control class would be led by a teacher without any training. By examining the gender differences in mathematics performance between schools/classes with a quasi-experimental variable, one would be better equipped to derive the causal factors underlying the differences.

Appendix A. – R Output for Exploratory IRT Analysis

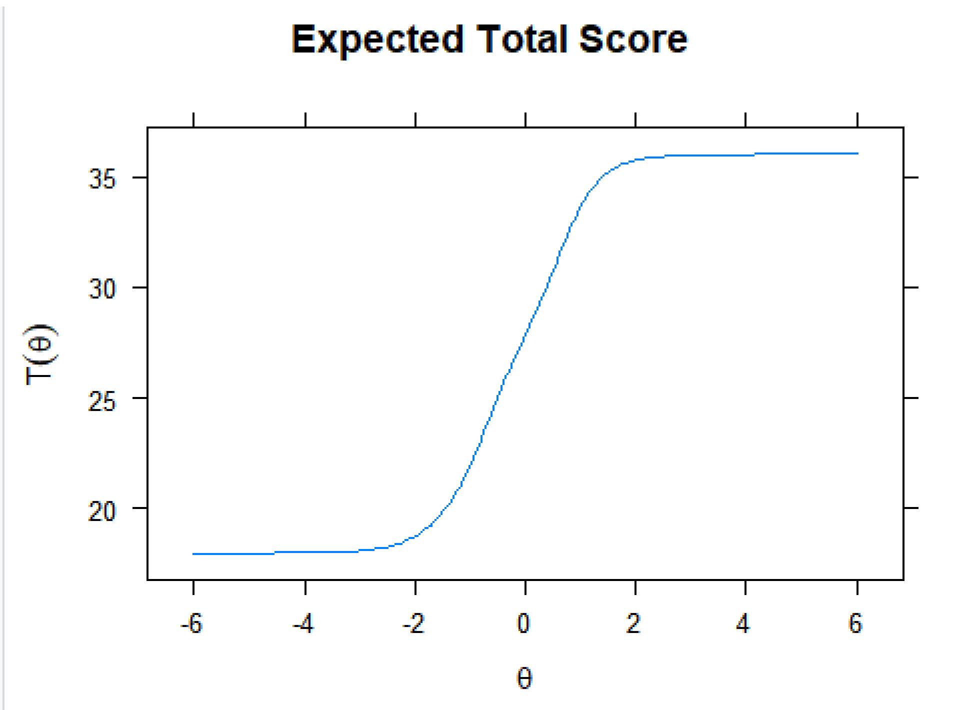
Teacher Confidence in Teaching Mathematics (TCONF) – TIMSS 2015 Regular

```
Iteration: 29, Log-Lik: -2517.897, Max-Change: 0.00010
> coef(mod2)
$rBTBM17A
       a1    d1     d2
par 2.697 2.194 -1.682

$rBTBM17B
       a1    d1     d2
par 2.342 3.454 -0.389

$rBTBM17C
       a1    d1     d2
par 1.651 1.585 -1.352

$rBTBM17D
      a1    d1     d2
par 3.27 2.206 -2.225

$rBTBM17E
       a1    d1     d2
par 3.487 2.506 -2.223

$rBTBM17F
       a1    d1     d2
par 2.332 3.396 -0.613

$rBTBM17G
       a1    d1     d2
par 2.703 2.233 -1.576

$rBTBM17H
       a1   d1     d2
par 3.136 1.99 -2.372

$rBTBM17I
       a1    d1     d2
par 3.275 2.228 -2.469

$GroupPars
    MEAN_1 COV_11
par      0      1

> summary(mod2, rotate = 'varimax')
            F1    h2
rBTBM17A 0.846 0.715
rBTBM17B 0.809 0.654
rBTBM17C 0.696 0.485
rBTBM17D 0.887 0.787
rBTBM17E 0.899 0.808
rBTBM17F 0.808 0.653
rBTBM17G 0.846 0.716
rBTBM17H 0.879 0.772
rBTBM17I 0.887 0.787
```

```
SS loadings:  6.377
Proportion Var:  0.709

Factor correlations:

   F1
F1  1
```

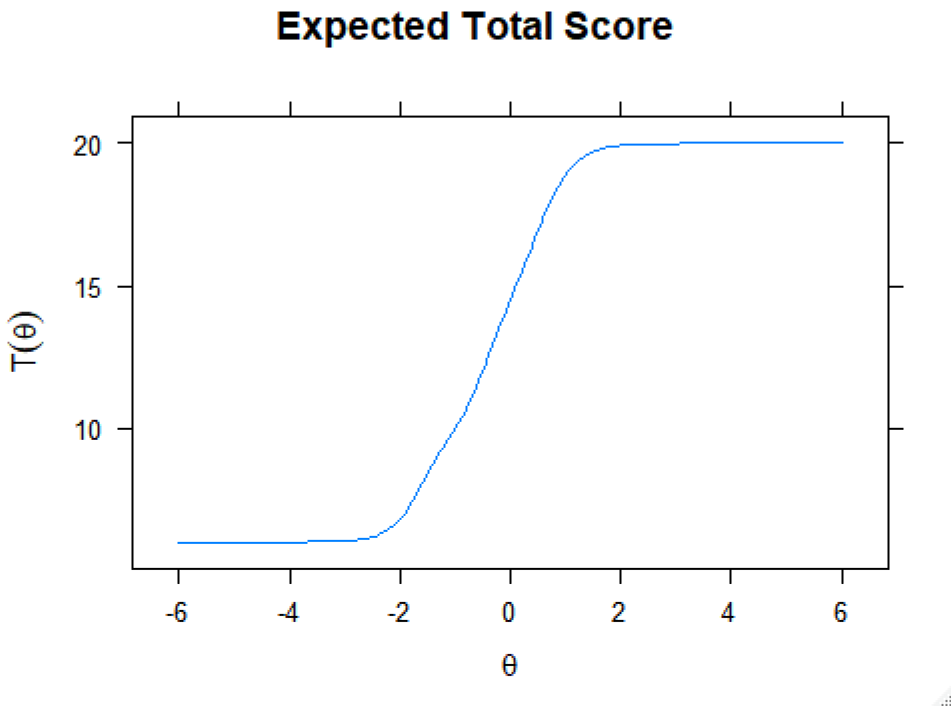
```
> residuals(mod2)
LD matrix (lower triangle) and standardized values:

          rBTBM17A rBTBM17B rBTBM17C rBTBM17D rBTBM17E rBTBM17F rBTBM17G rBTBM17H rBTBM17I
rBTBM17A        NA   -0.106   -0.129    0.107    0.142   -0.131   -0.100   -0.133   -0.058
rBTBM17B    -8.232       NA    0.145   -0.071   -0.080    0.102   -0.144   -0.114   -0.085
rBTBM17C   -12.270   15.426       NA   -0.042   -0.094    0.073   -0.119   -0.153    0.121
rBTBM17D     8.325   -3.724   -1.311       NA    0.112   -0.172   -0.152    0.119   -0.133
rBTBM17E    14.805   -4.697   -6.449    9.246       NA   -0.161   -0.100   -0.151   -0.153
rBTBM17F   -12.475    7.615    3.916  -21.702  -19.018       NA    0.188   -0.099    0.103
rBTBM17G    -7.331  -15.328  -10.490  -16.884   -7.338   25.894       NA   -0.144    0.133
rBTBM17H   -13.036   -9.568  -17.142   10.335  -16.677   -7.176  -15.177       NA   -0.110
rBTBM17I    -2.442   -5.343   10.658  -12.983  -17.103    7.718   12.910   -8.778       NA
```

```
> plot(mod2)
```



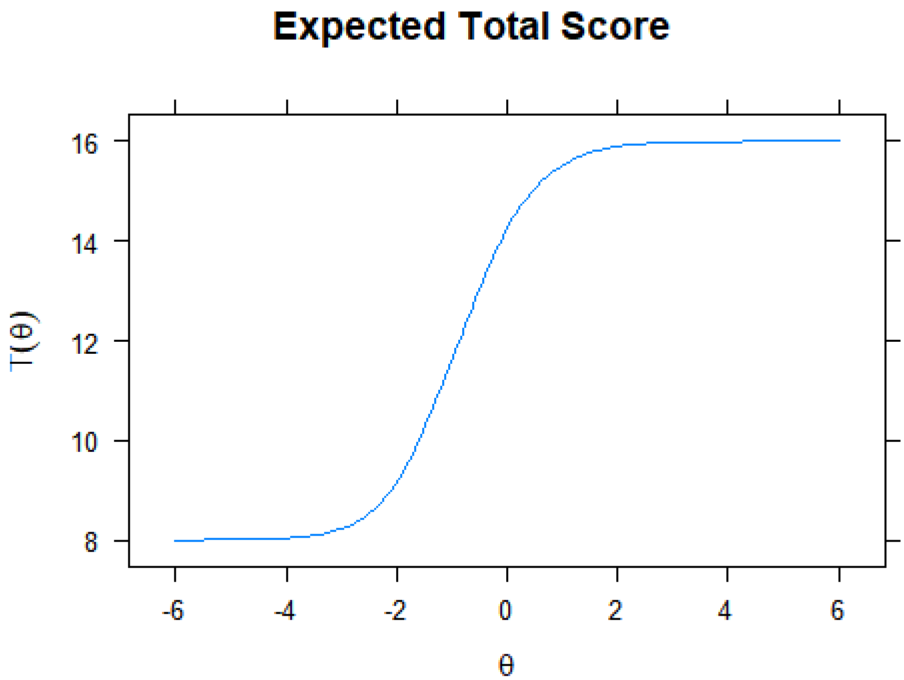
Communication Among Teachers (TCOM) – TIMSS 2015 Regular

```
Iteration: 52, Log-Lik: -1635.825, Max-Change: 0.00010
> coef(mod2)
$rBTBG09A
      a1    d1     d2
par 4.04 2.037 -2.089

$rBTBG09B
       a1    d1    d2     d3
par 4.809 7.992 2.373 -1.668

$rBTBG09C
      a1    d1    d2     d3
par 3.36 6.107 1.495 -2.166

$rBTBG09E
       a1    d1    d2     d3
par 2.836 3.809 0.233 -2.814

$rBTBG09F
       a1    d1    d2     d3
par 3.778 5.728 1.417 -2.146

$GroupPars
    MEAN_1 COV_11
par      0      1

> summary(mod2, rotate = 'varimax')
            F1    h2
rBTBG09A 0.922 0.849
rBTBG09B 0.943 0.889
rBTBG09C 0.892 0.796
rBTBG09E 0.857 0.735
rBTBG09F 0.912 0.831

SS loadings:  4.1
Proportion Var:  0.82

Factor correlations:

   F1
F1  1

> residuals(mod2)
LD matrix (lower triangle) and standardized values:

         rBTBG09A rBTBG09B rBTBG09C rBTBG09E rBTBG09F
rBTBG09A       NA    0.139    0.118   -0.169   -0.119
rBTBG09B   14.232       NA   -0.252   -0.232    0.149
rBTBG09C   10.266  -70.443       NA    0.125   -0.224
rBTBG09E  -21.047  -59.372   17.292       NA    0.194
rBTBG09F  -10.531   24.482  -55.392   41.487       NA
```


Expected Total Score
T(θ)
20
15
10
-6
-4
-2
0
2
4
6
θ

Teacher Support of Student Participation (TSUP) – TIMSS 2015 Regular

```
> mod2 <- mirt(data14, 1)
Iteration: 22, Log-Lik: -1176.054, Max-Change: 0.00006
> coef(mod2)
$rBTBG14B
       a1    d1    d2
par 1.945 3.991 1.576

$rBTBG14D
       a1    d1    d2
par 2.099 2.702 0.389

$rBTBG14F
       a1    d1     d2
par 1.713 1.678 -0.995

$rBTBG14G
       a1   d1    d2
par 3.079 4.51 1.463

$GroupPars
    MEAN_1 COV_11
par      0      1

> summary(mod2, rotate = 'varimax')
            F1    h2
rBTBG14B 0.753 0.566
rBTBG14D 0.777 0.603
rBTBG14F 0.709 0.503
rBTBG14G 0.875 0.766

SS loadings:  2.439
Proportion Var:  0.61

Factor correlations:

   F1
F1  1

> residuals(mod2)
LD matrix (lower triangle) and standardized values:

         rBTBG14B rBTBG14D rBTBG14F rBTBG14G
rBTBG14B       NA    0.090   -0.108   -0.079
rBTBG14D    5.956       NA    0.047   -0.137
rBTBG14F   -8.558    1.610       NA    0.068
rBTBG14G   -4.628  -13.781    3.443       NA


> plot(mod2)
```

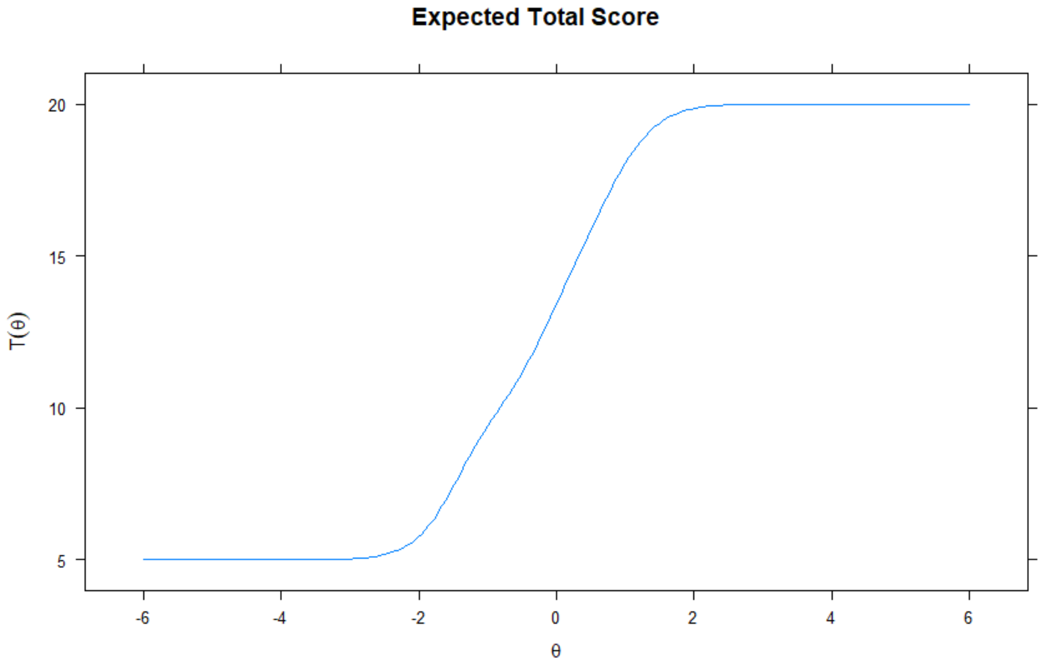


Expected Total Score
16
14
12
10
8
T(θ)
-6
-4
-2
0
2
4
6
θ

Communication among Teachers (TCOM) – TIMSS 2015 Advanced

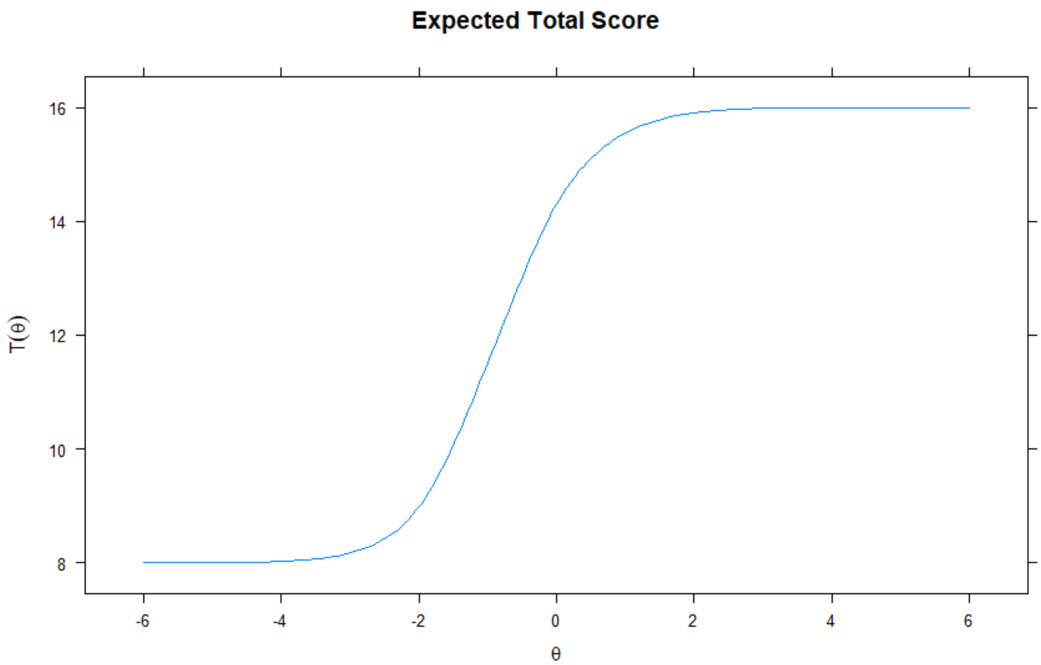
```
> mod2 <- mirt(dataA9, 1)
Iteration: 36, Log-Lik: -1488.357, Max-Change: 0.00009
> coef(mod2)
$`rMTBG09A`
       a1    d1    d2     d3
par 3.371 5.091 0.805 -2.445

$rMTBG09B
       a1    d1    d2     d3
par 3.576 5.289 0.713 -1.859

$rMTBG09C
       a1    d1    d2     d3
par 3.366 5.248 0.659 -2.822

$rMTBG09E
       a1  d1     d2     d3
par 3.295 3.9 -0.706 -4.161

$rMTBG09F
       a1    d1    d2     d3
par 2.943 4.522 1.032 -2.131

$GroupPars
    MEAN_1 COV_11
par      0      1

> summary(mod2, rotate = 'varimax')
             F1    h2
rMTBG09A 0.893 0.797
rMTBG09B 0.903 0.815
rMTBG09C 0.892 0.796
rMTBG09E 0.888 0.789
rMTBG09F 0.866 0.749

SS loadings:  3.947
Proportion Var:  0.789

Factor correlations:

   F1
F1  1

> residuals(mod2)
LD matrix (lower triangle) and standardized values:

         rMTBG09A rMTBG09B rMTBG09C rMTBG09E rMTBG09F
rMTBG09A       NA    0.274    0.154   -0.142   -0.247
rMTBG09B   69.570       NA   -0.205   -0.246    0.142
rMTBG09C   21.970  -38.981       NA   -0.152   -0.149
rMTBG09E  -18.627  -56.233  -21.500       NA    0.161
rMTBG09F  -56.484   18.655  -20.596   24.120       NA
```

```
> plot(mod2)
```



Teacher Support of Student Participation (TSUP) – TIMSS 2015 Advanced

```
> mod2 <- mirt(dataA14, 1)
Iteration: 23, Log-Lik: -996.956, Max-Change: 0.00007
> coef(mod2)
$`rMTBG14B`
       a1    d1    d2
par 1.721 3.036 0.973

$rMTBG14D
       a1    d1   d2
par 2.805 3.754 1.04

$rMTBG14F
       a1    d1     d2
par 1.749 1.662 -0.948

$rMTBG14G
      a1    d1    d2
par 2.83 4.212 1.374

$GroupPars
    MEAN_1 COV_11
par      0      1

> summary(mod2, rotate = 'varimax')
            F1    h2
rMTBG14B 0.711 0.506
rMTBG14D 0.855 0.731
rMTBG14F 0.717 0.514
rMTBG14G 0.857 0.734

SS loadings:  2.485
Proportion Var:  0.621

Factor correlations:

   F1
F1  1
> residuals(mod2)
LD matrix (lower triangle) and standardized values:

         rMTBG14B rMTBG14D rMTBG14F rMTBG14G
rMTBG14B       NA    0.097   -0.063   -0.102
rMTBG14D    5.758       NA   -0.069   -0.081
rMTBG14F   -2.429   -2.947       NA    0.136
rMTBG14G   -6.451   -4.006   11.365       NA
> plot(mod2)
```


Expected Total Score
T(θ)
16
14
12
10
8
-6
-4
-2
0
2
4
6
θ

Appendix B. – TIMSS 2015 Regular

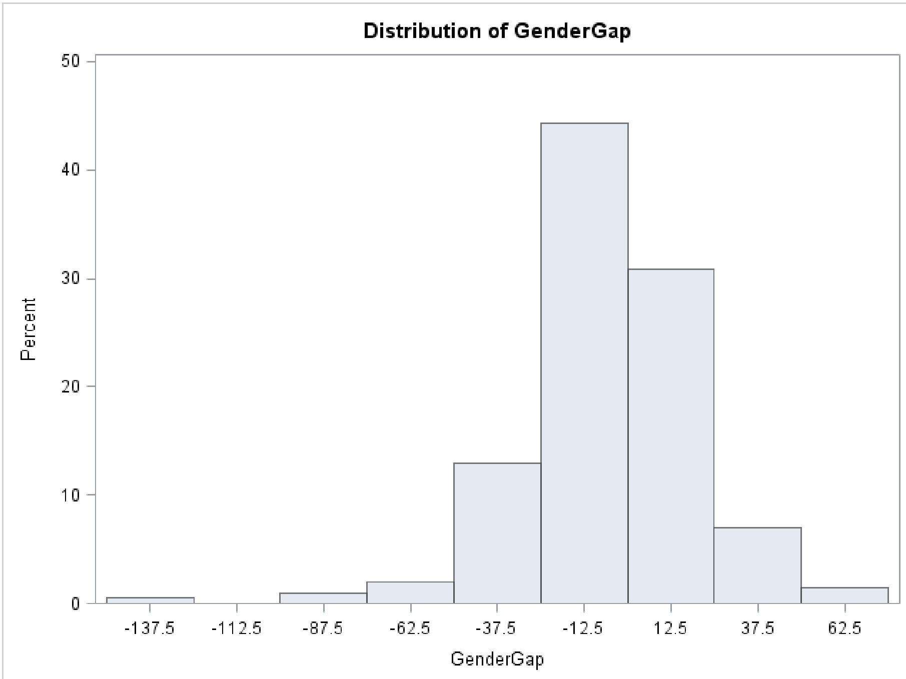


Figure B.1. Distribution of within-school gender gap across schools

Table B.1. Percentage of missingness in each variable (TIMSS 2015 Regular)

Variable Name	% of missingness
Instruction affected by mathematics resources shortage (BCBGMRS)	10.2%
School mean SES (MEAN_SES)	0.0%
Teacher confidence in teaching mathematics (TCONF)	10.7%
Communication among teachers (TCOM)	10.2%
Teacher support of student participation (TSUP)	10.2%
Shortage of teachers (SHORTT)	11.9%

Distribution of Level 2 Variables (TIMSS 2015 Regular)

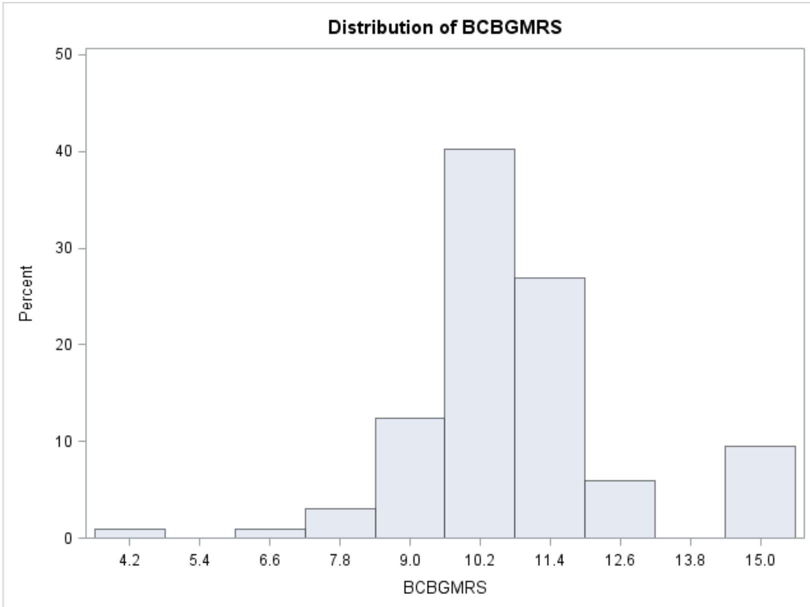


Figure B.2. Distribution of BCBGMRS (Instruction affected by mathematics resources shortage) variable (TIMSS 2015 Regular)

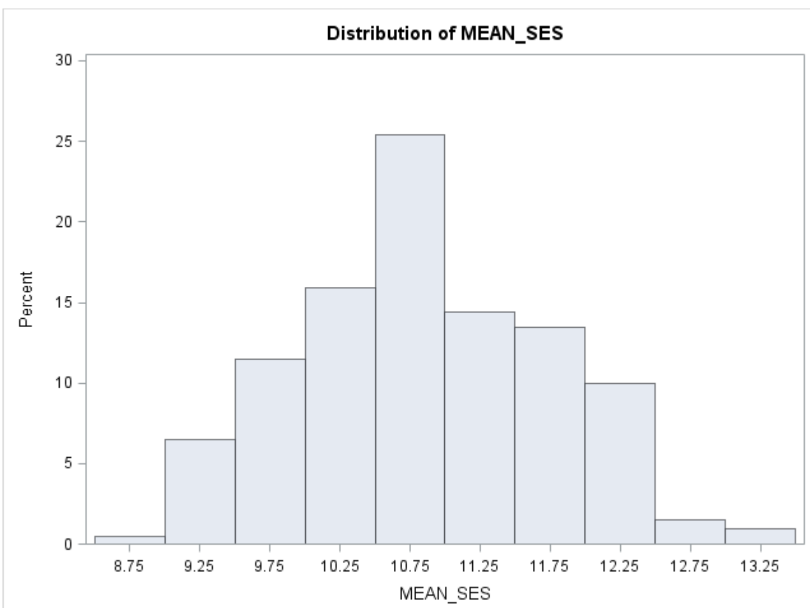


Figure B.3. Distribution of school mean SES variable (TIMSS 2015 Regular)

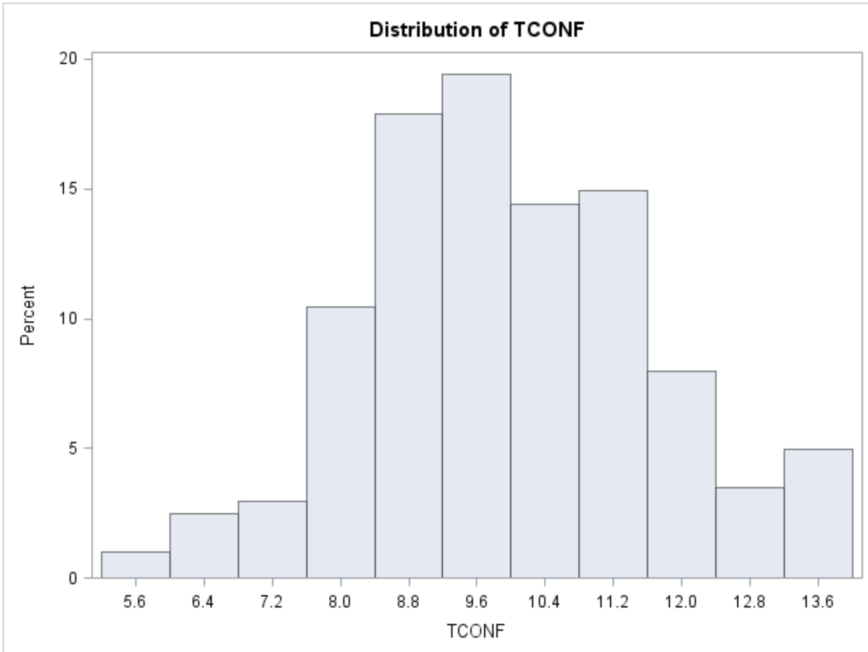


Figure B.4. Distribution of TCONF (teacher confidence in teaching mathematics) (TIMSS 2015 Regular)

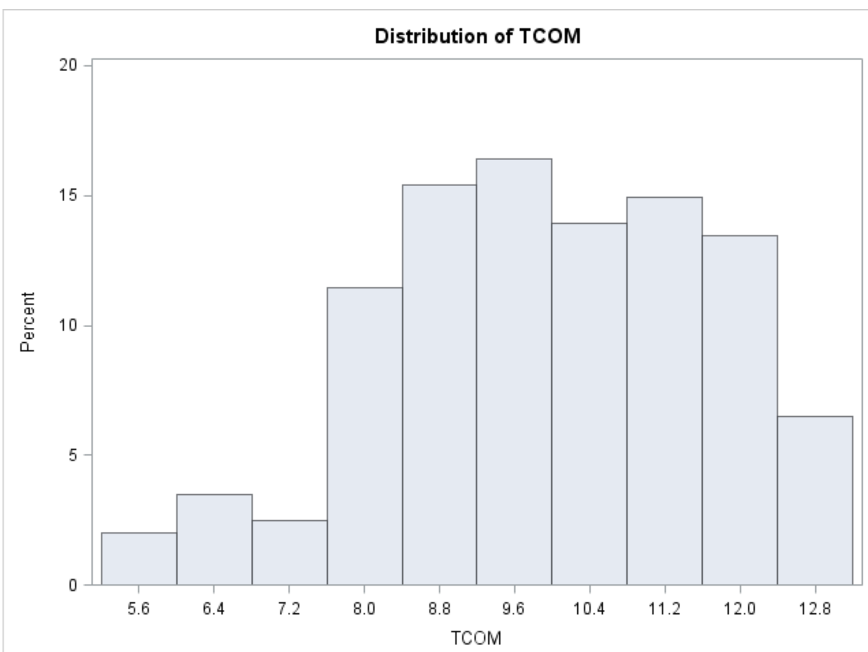


Figure B.5. Distribution of TCOM (communication among teachers) variable (TIMSS 2015 Regular)

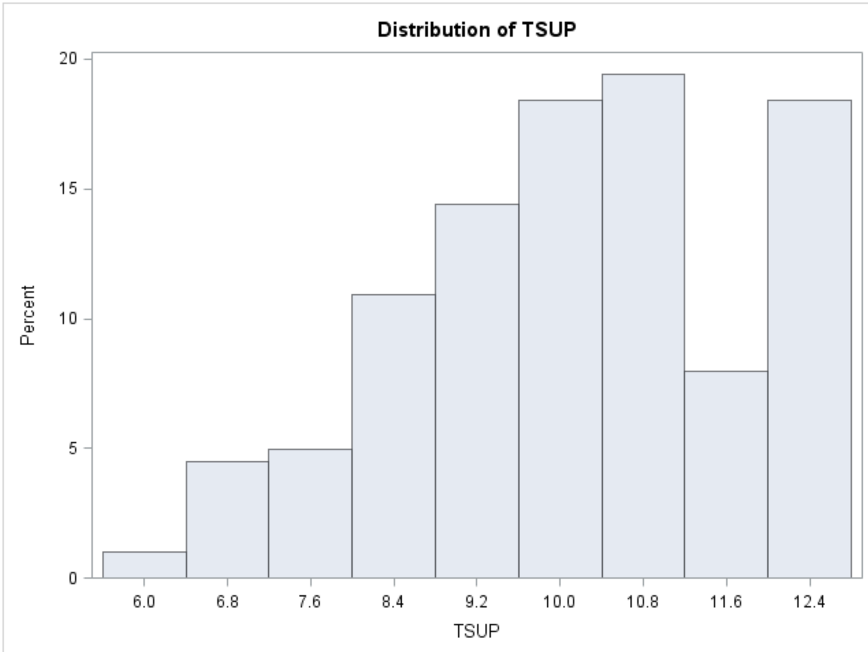


Figure B.6. Distribution of TSUP (teacher support of student participation) variable (TIMSS 2015 Regular)

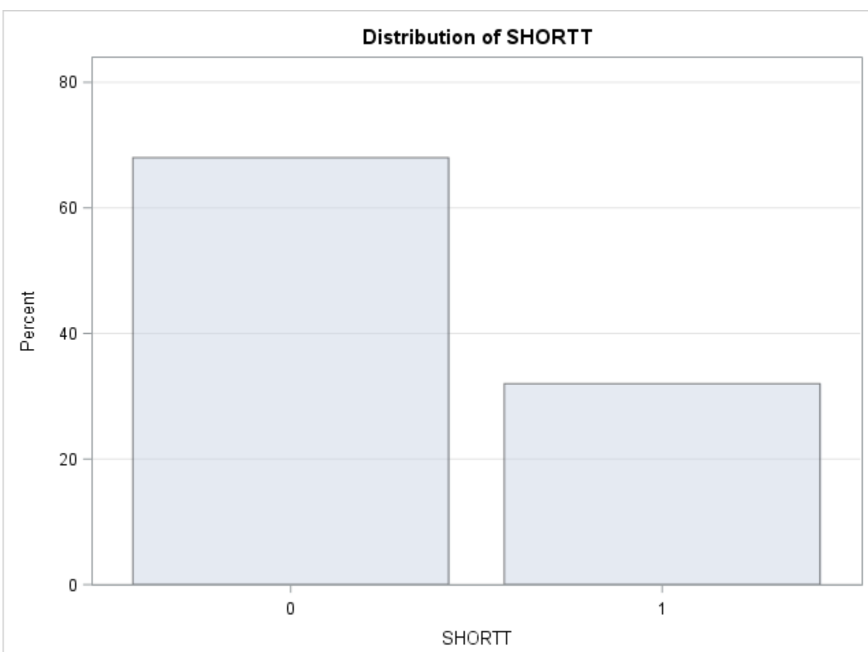


Figure B.7. Distribution of teacher shortage (SHORTT) variable (TIMSS 2015 Regular)

Scatter plots between Level 2 Variables (TIMSS 2015 Regular)

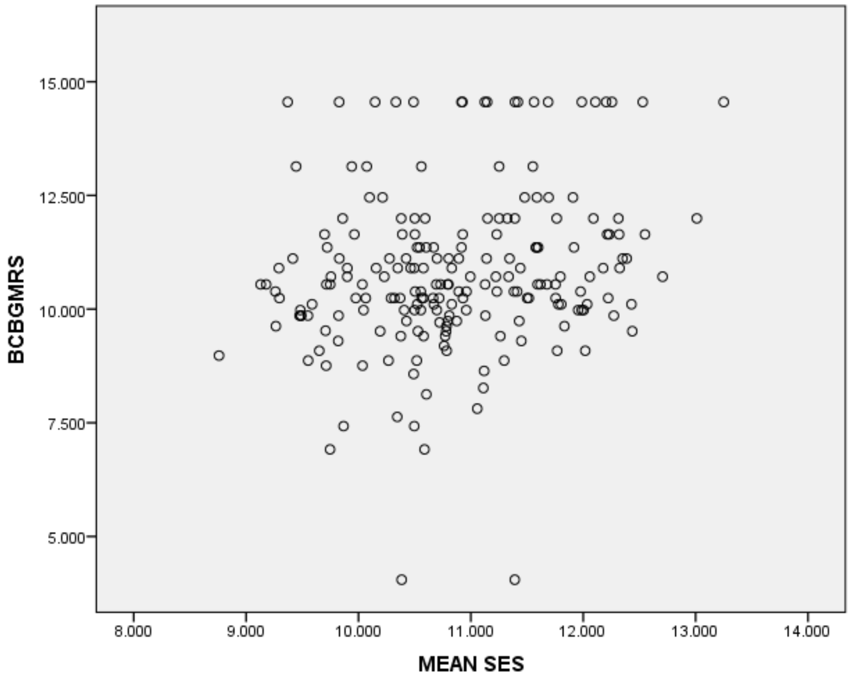


Figure B.8. Scatterplot of BCBGMRS and MEAN SES (TIMSS 2015 Regular)

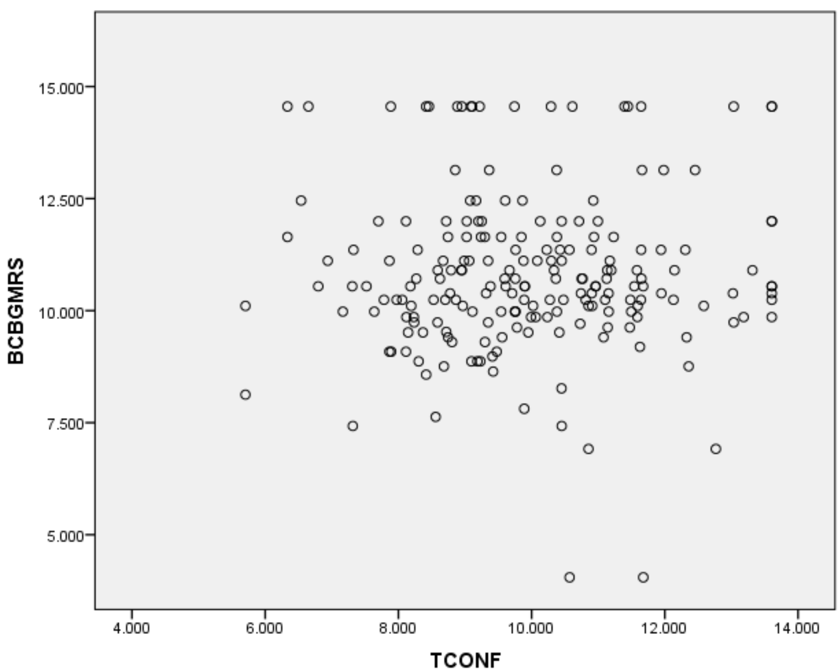


Figure B.9. Scatterplot of BCBGMRS and TCONF (TIMSS 2015 Regular)

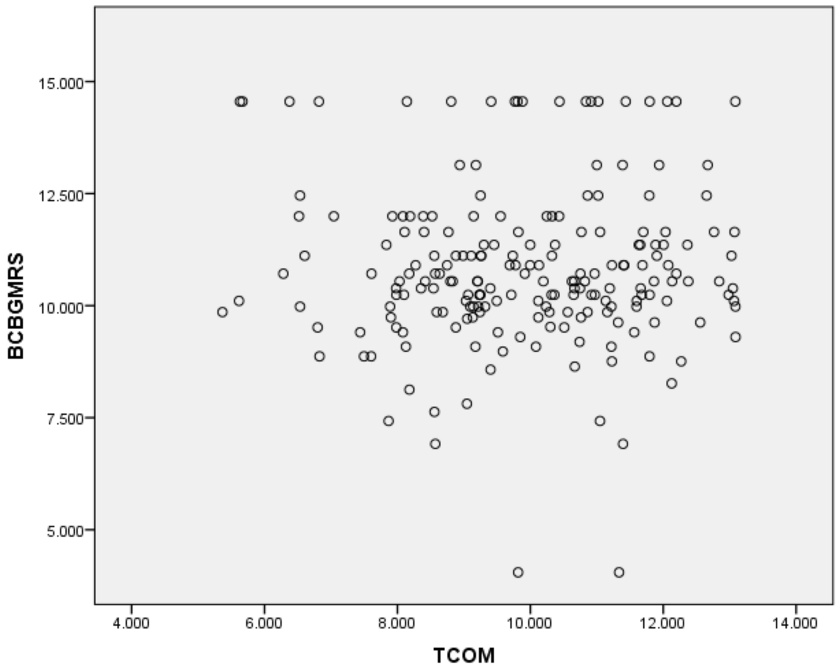


Figure B.10. Scatterplot of BCBGMRS and TCOM (TIMSS 2015 Regular)

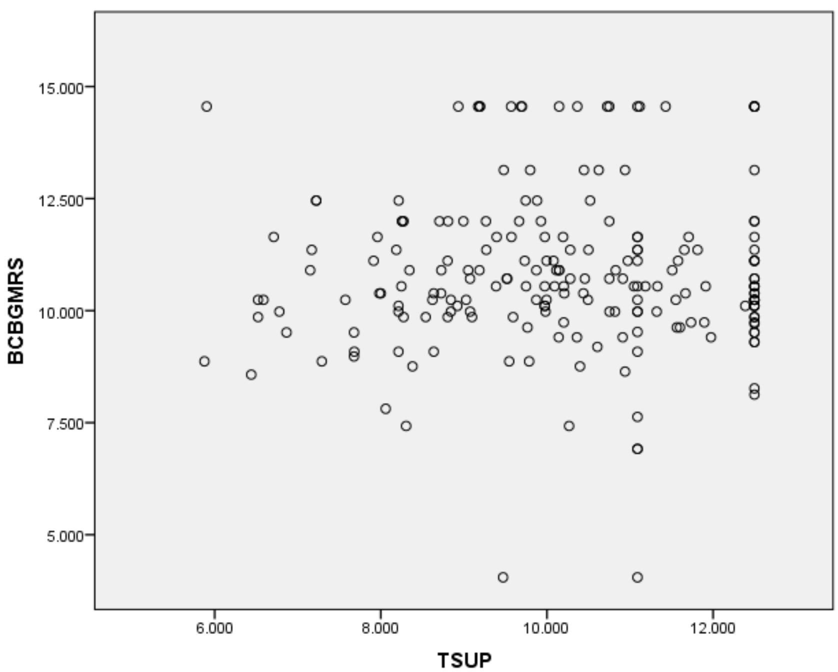


Figure B.11. Scatterplot of BCBGMRS and TSUP (TIMSS 2015 Regular)

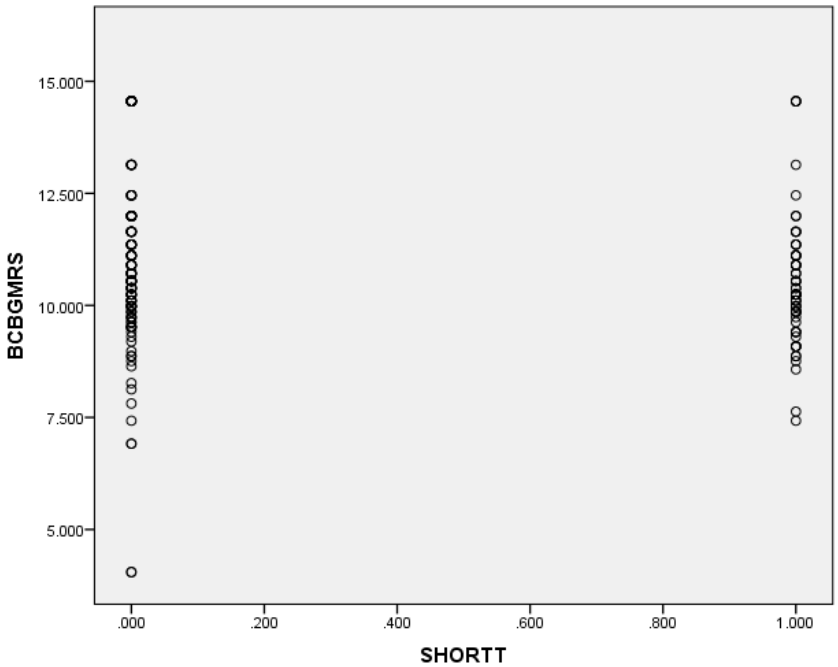


Figure B.12. Scatterplot of BCBGMRS and SHORTT (TIMSS 2015 Regular)

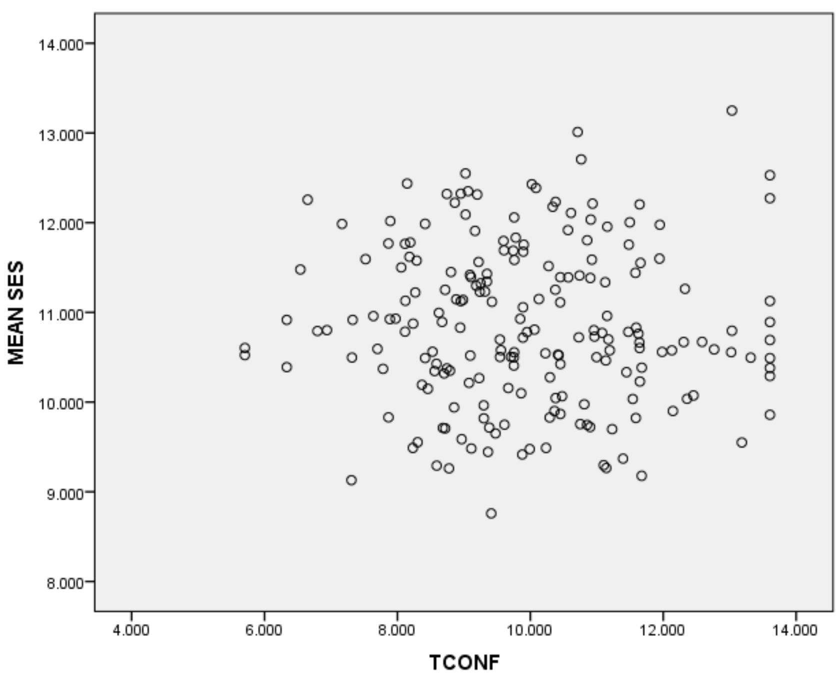


Figure B.13. Scatterplot of MEAN SES and TCONF (TIMSS 2015 Regular)

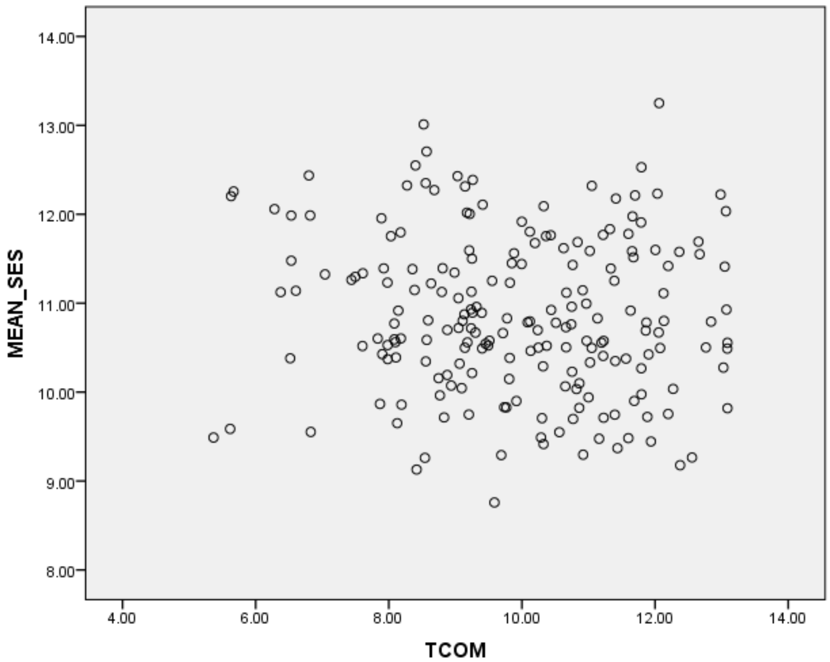


Figure B.14. Scatterplot of MEAN SES and TCOM (TIMSS 2015 Regular)

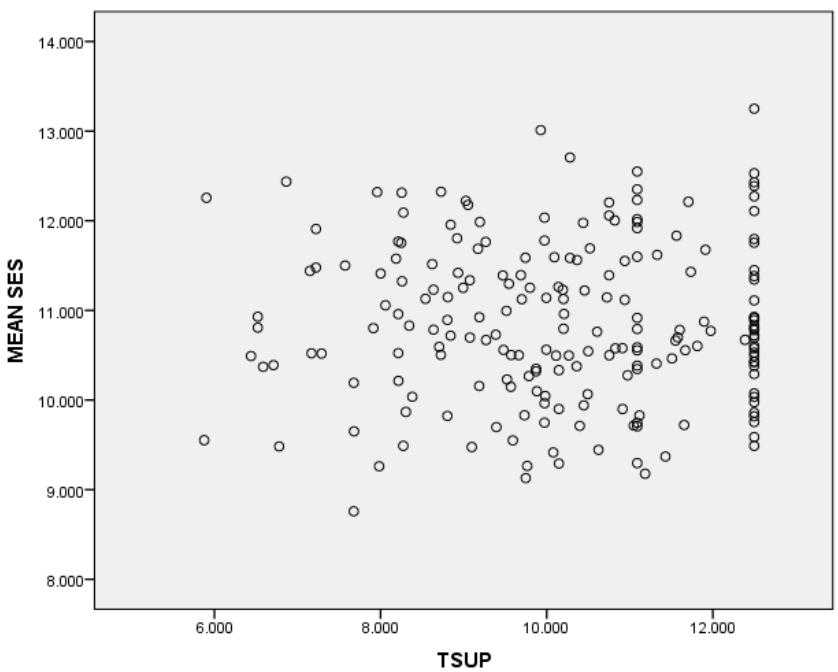


Figure B.15. Scatterplot of MEAN SES and TSUP (TIMSS 2015 Regular)

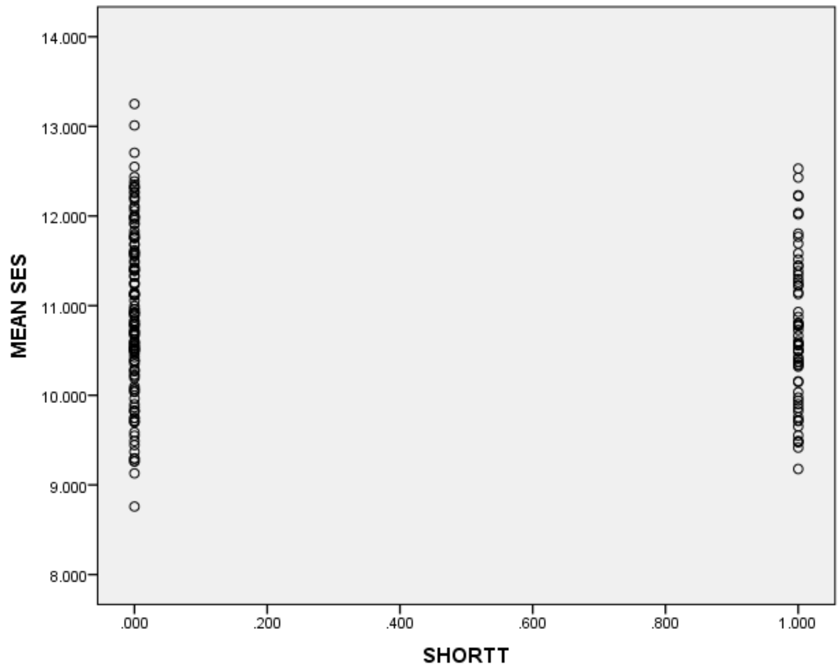


Figure B.16. Scatterplot of MEAN SES and SHORTT (TIMSS 2015 Regular)

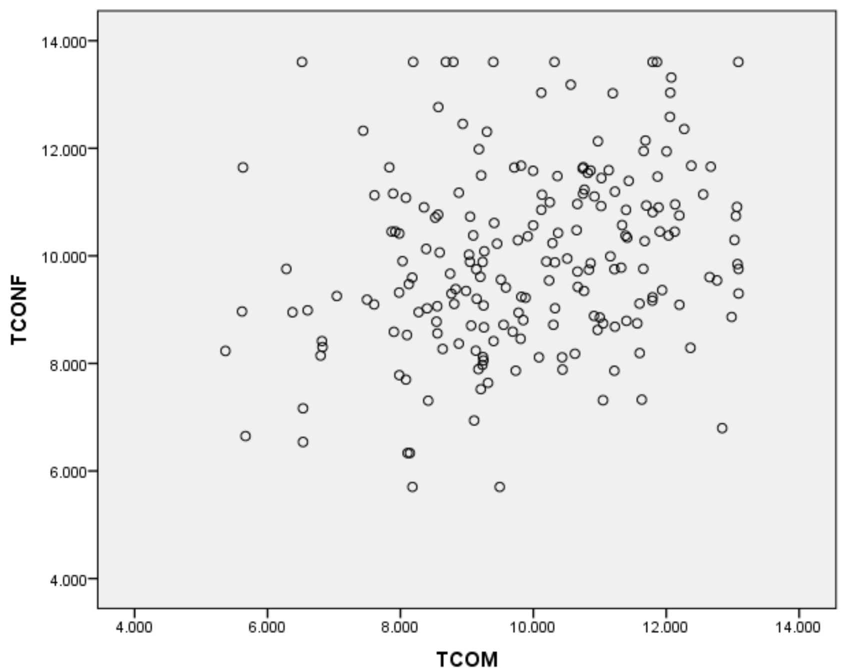


Figure B.17. Scatterplot of TCONF and TCOM (TIMSS 2015 Regular)

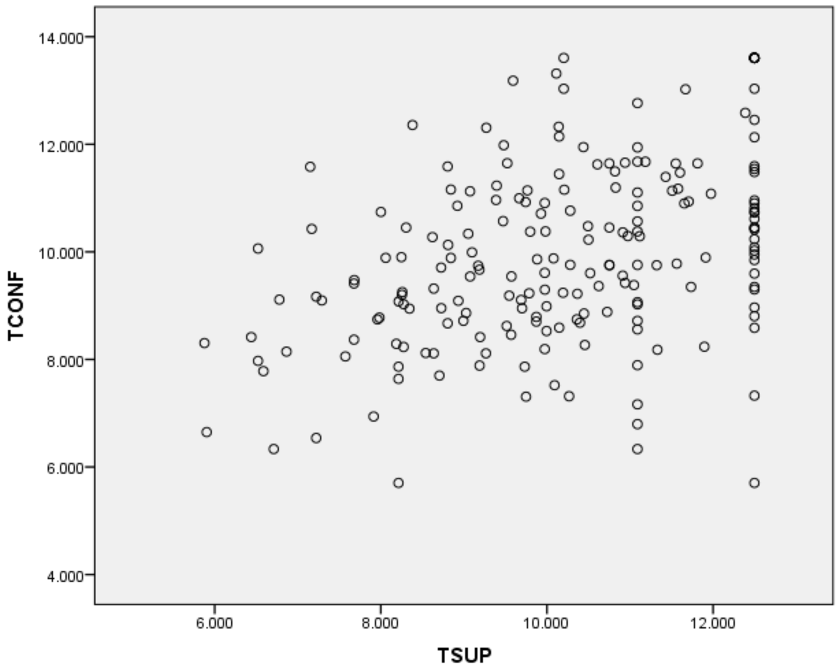


Figure B.18. Scatterplot of TCONF and TSUP (TIMSS 2015 Regular)

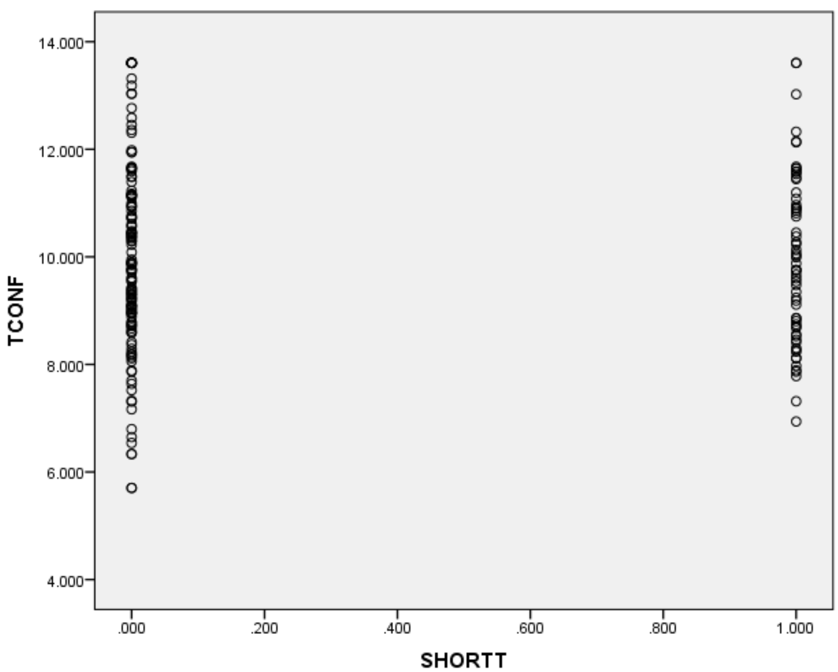


Figure B.19. Scatterplot of TCONF and SHORTT (TIMSS 2015 Regular)

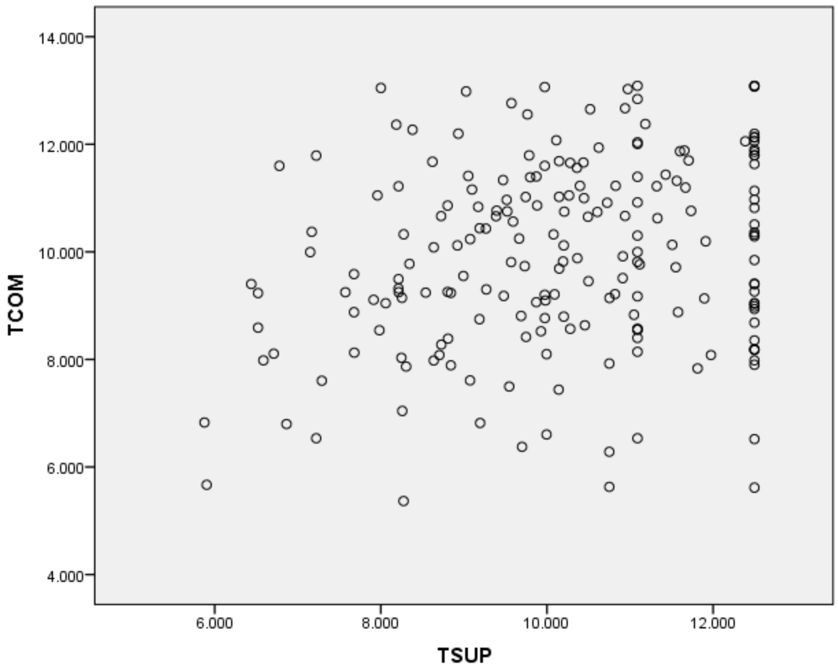


Figure B.20. Scatterplot of TCOM and TSUP (TIMSS 2015 Regular)

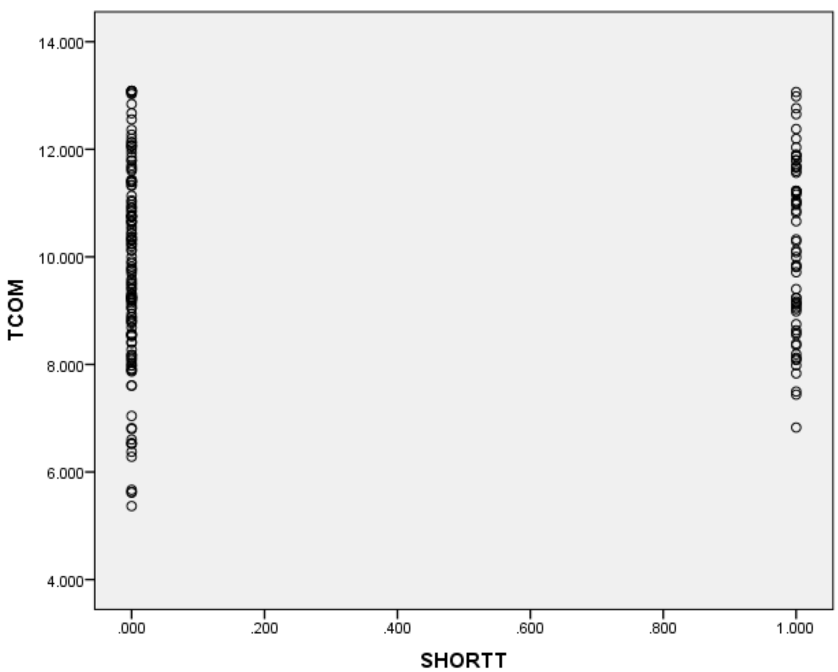


Figure B.21. Scatterplot of TCOM and SHORTT (TIMSS 2015 Regular)

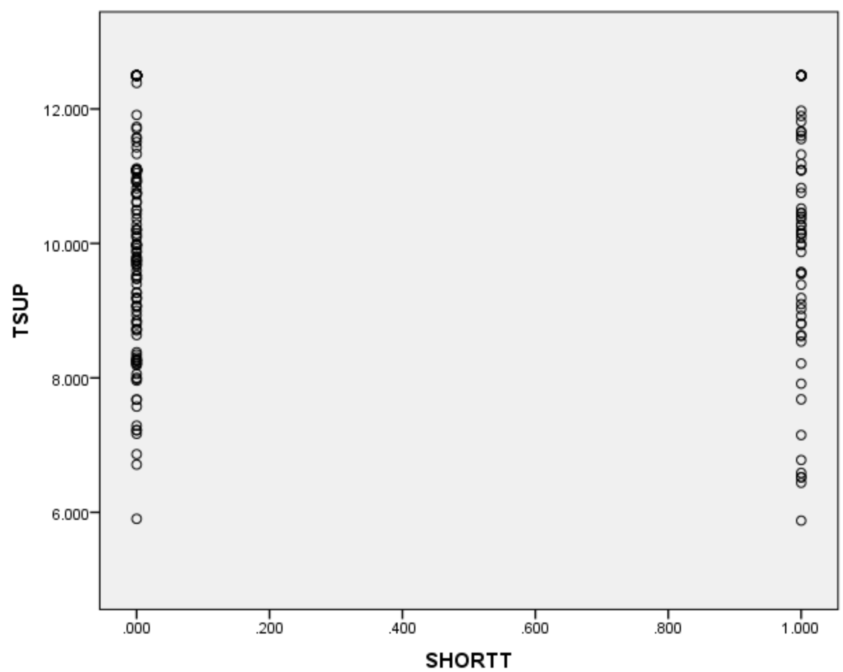


Figure B.22. Scatterplot of TSUP and SHORTT (TIMSS 2015 Regular)

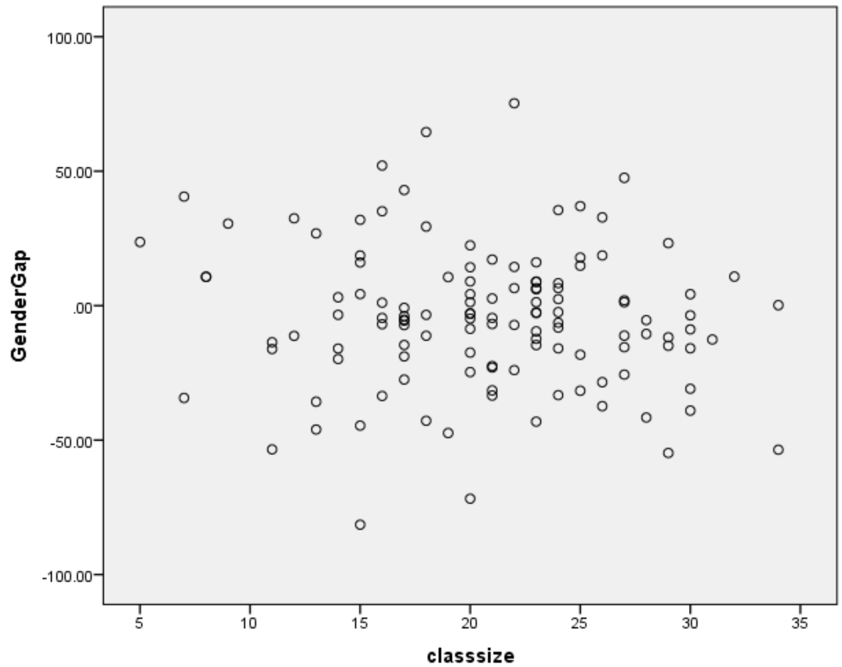


Figure B.23. Scatterplot of gender gap and class size (N=124)

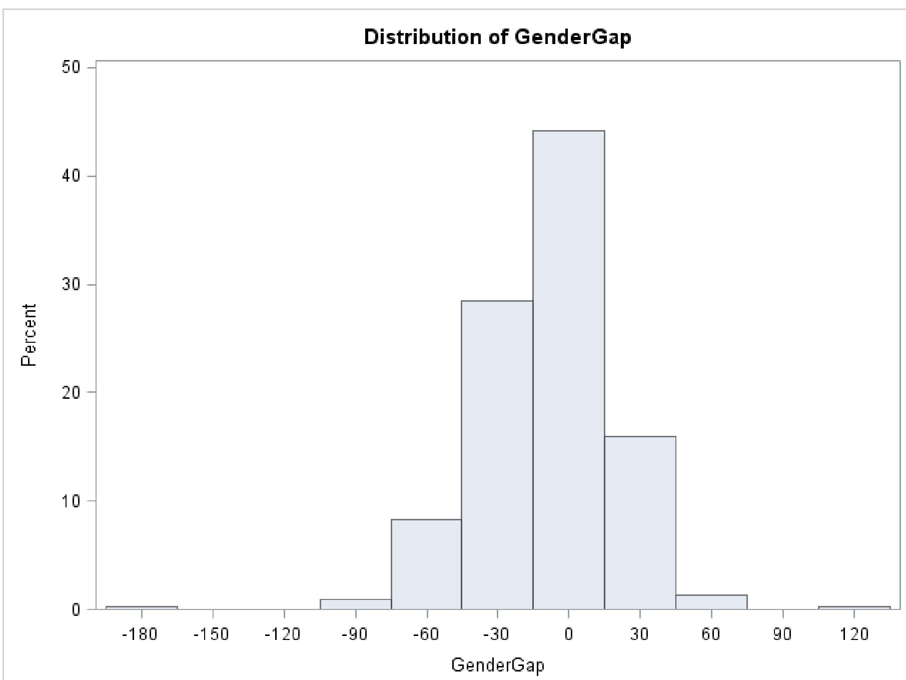


Figure B.24. Distribution of within-school gender gap across teachers

Appendix C. – TIMSS 2015 Advanced

Table C.1. Percentage of missingness in each variable (TIMSS 2015 Advanced)

Variable Name	% of missingness
Safe and orderly school (MTBGSOS)	15.3%
School mean SES (MEAN_SES)	0.0%
Communication among teachers (TCOM)	15.7%
Teacher support of student participation (TSUP)	15.7%
Shortage of teachers (SHORTT)	16.1%

Distribution of Level 2 Variables (TIMSS 2015 Advanced)

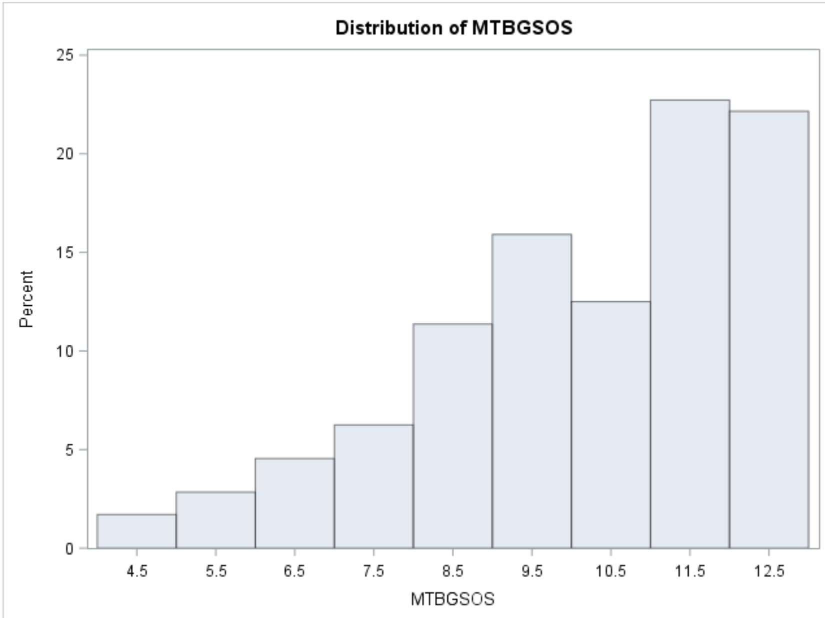


Figure C.1. Distribution of MTBGSOS (Safe and orderly school) variable (TIMSS 2015 Advanced)

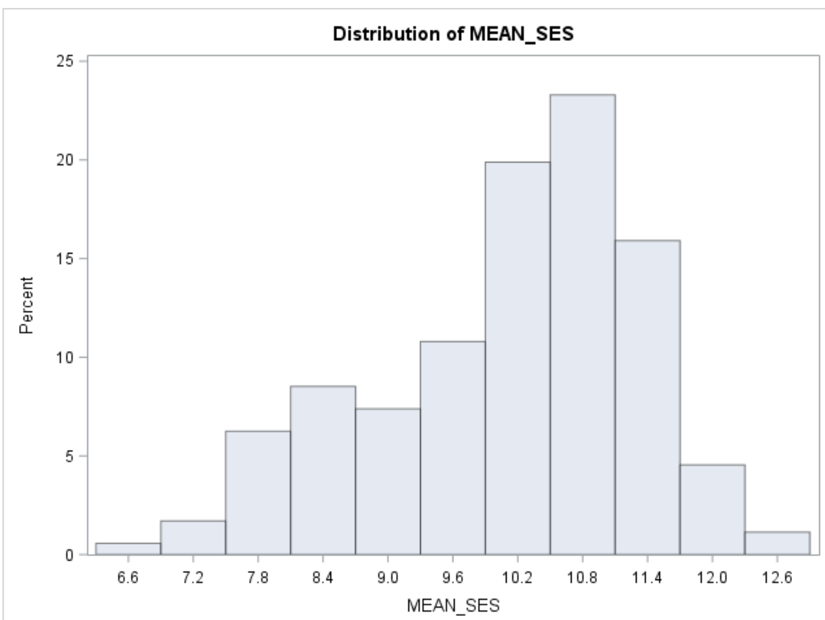


Figure C.2. Distribution of MEAN SES variable (TIMSS 2015 Advanced)

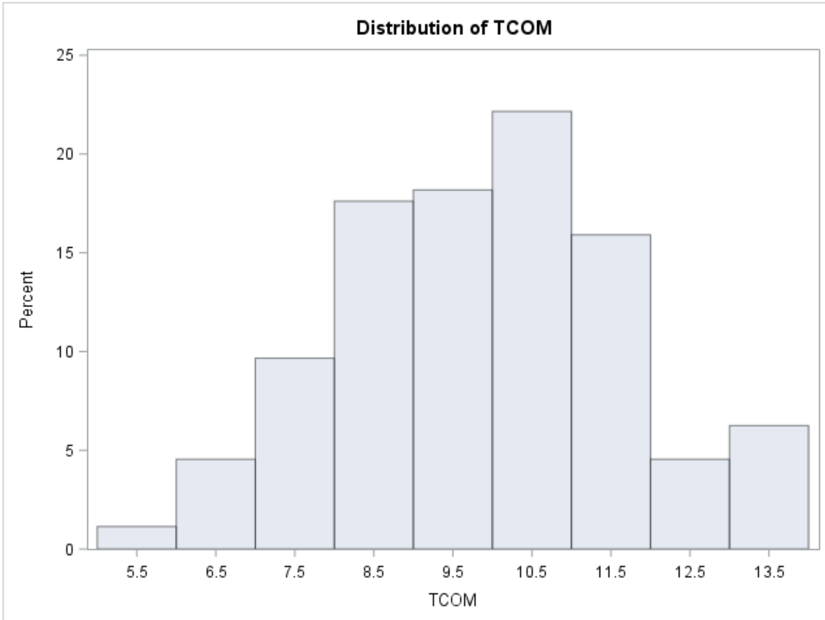


Figure C.3. Distribution of TCOM variable (TIMSS 2015 Advanced)

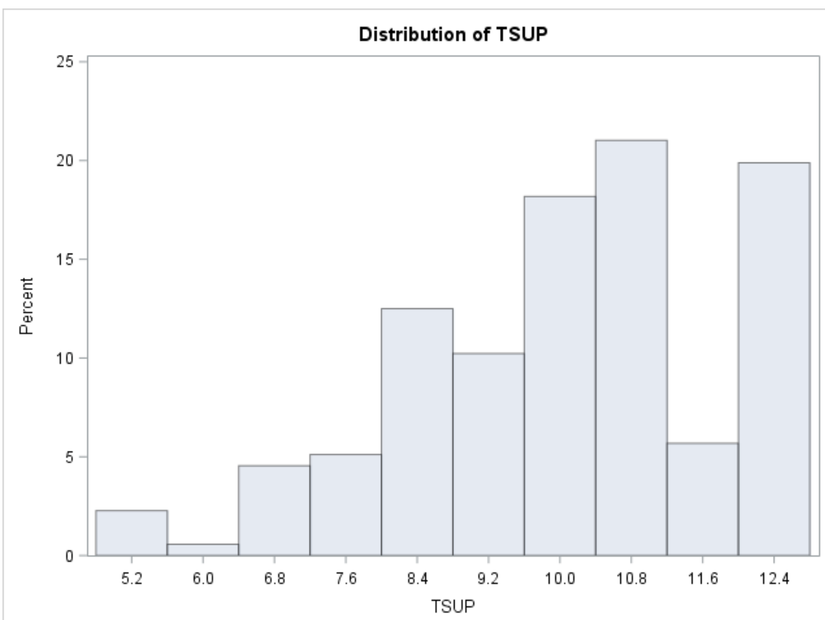


Figure C.4. Distribution of TSUP variable (TIMSS 2015 Advanced)

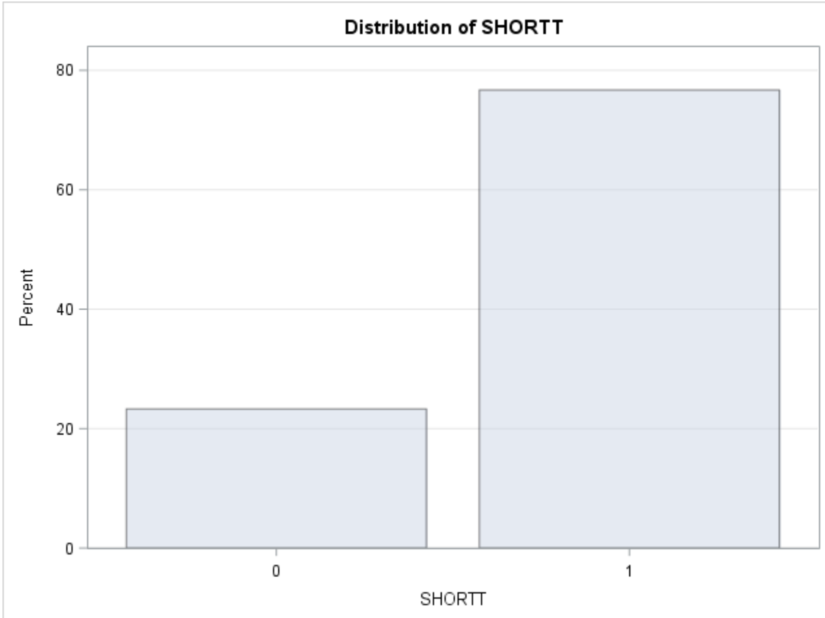


Figure C.5. Distribution of SHORTT variable (TIMSS 2015 Advanced)

Scatter plots between Level 2 Variables (TIMSS 2015 Advanced)

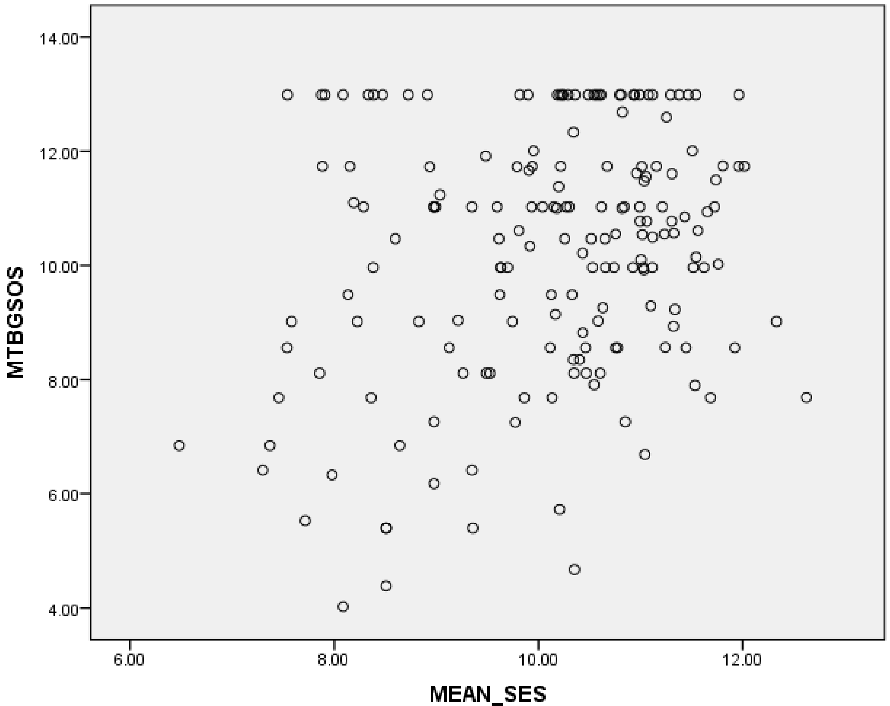


Figure C.6. Scatterplot of MTBGSOS and MEAN SES (TIMSS 2015 Advanced)

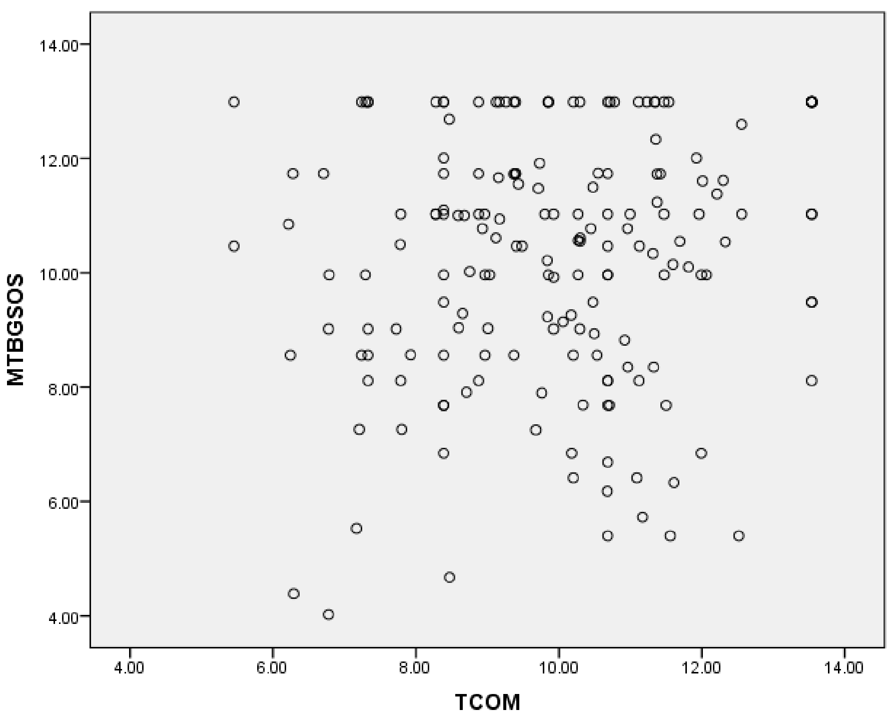


Figure C.7. Scatterplot of MTBGSOS and TCOM (TIMSS 2015 Advanced)

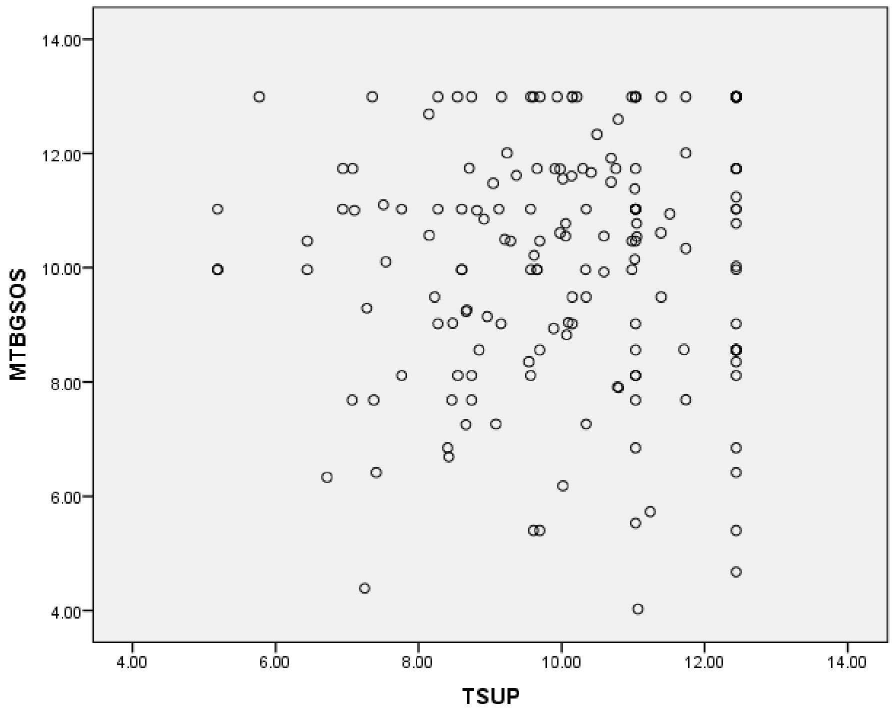


Figure C.8. Scatterplot of MTBGSOS and TSUP (TIMSS 2015 Advanced)

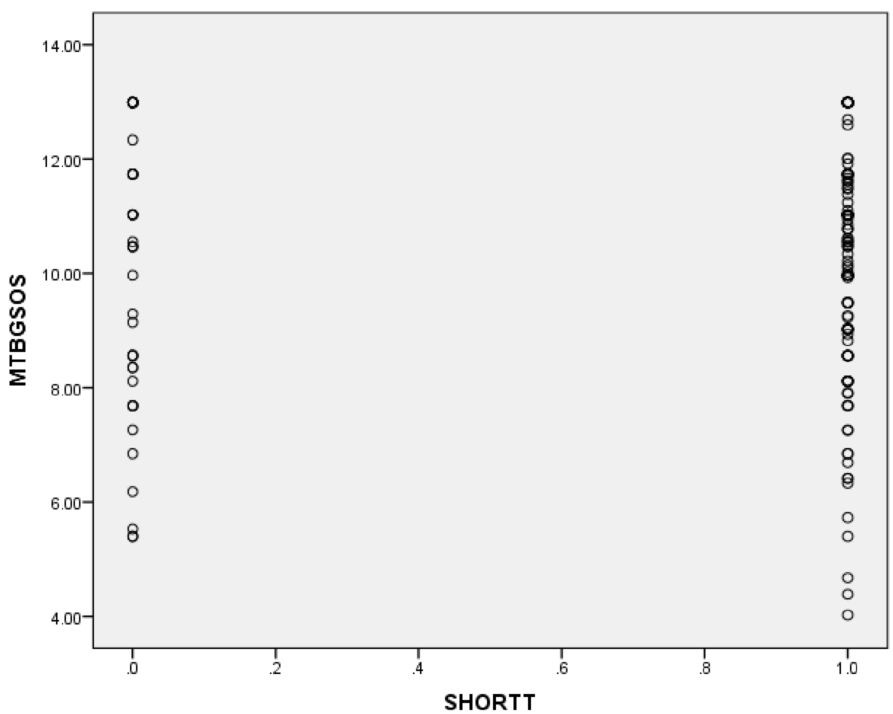


Figure C.9. Scatterplot of MTBGSOS and SHORTT (TIMSS 2015 Advanced)

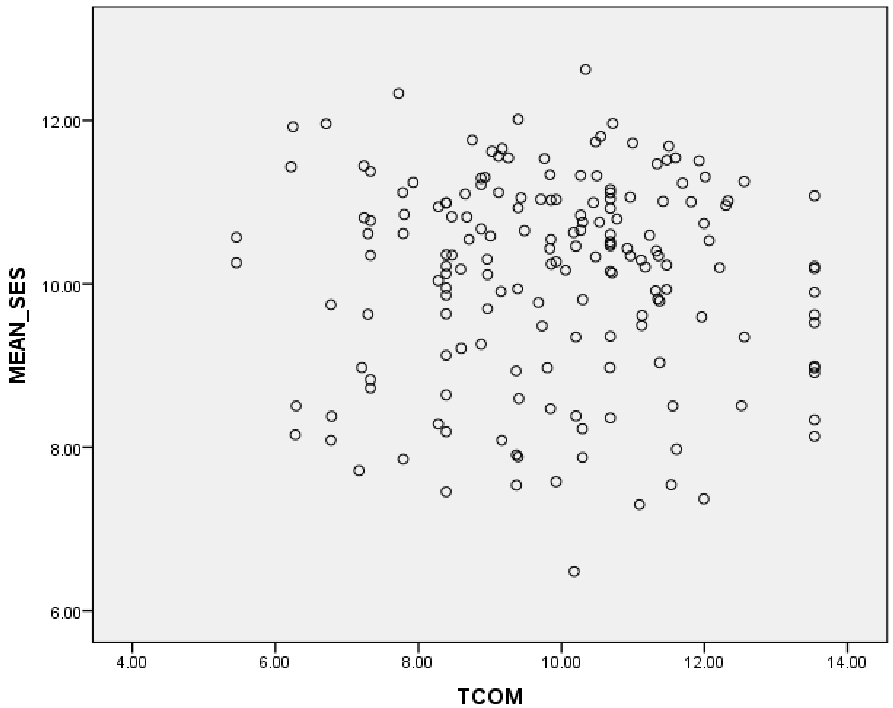


Figure C.10. Scatterplot of MEAN SES and TCOM (TIMSS 2015 Advanced)

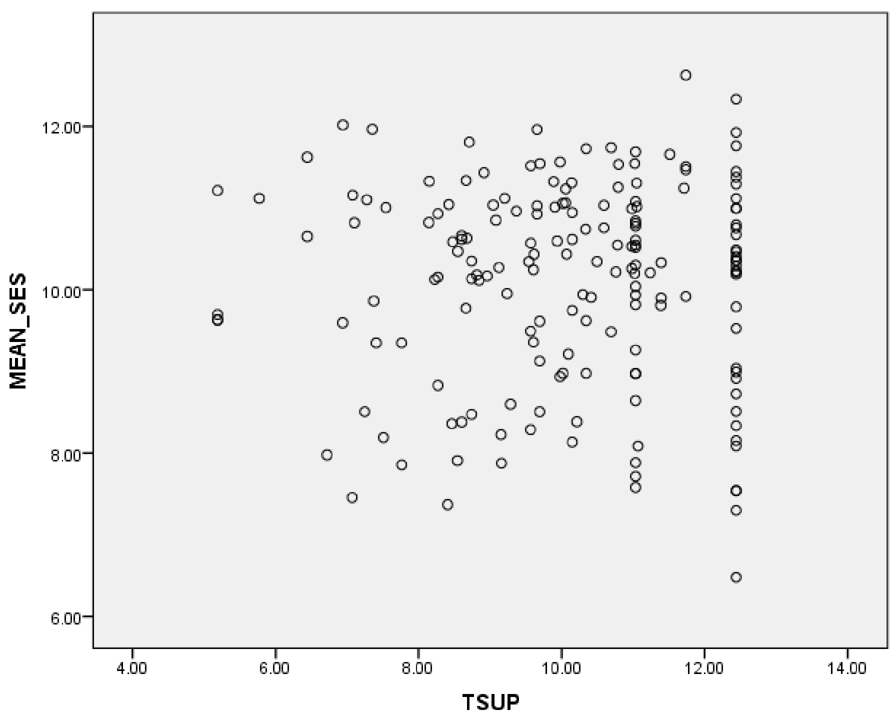


Figure C.11. Scatterplot of MEAN SES and TSUP (TIMSS 2015 Advanced)

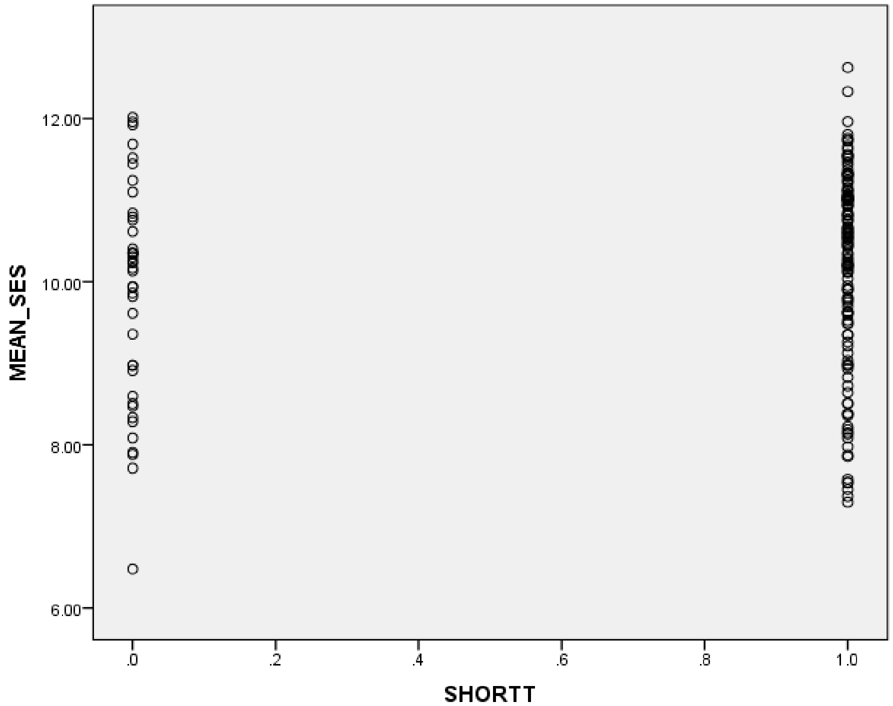


Figure C.12. Scatterplot of MEAN SES and SHORTT (TIMSS 2015 Advanced)

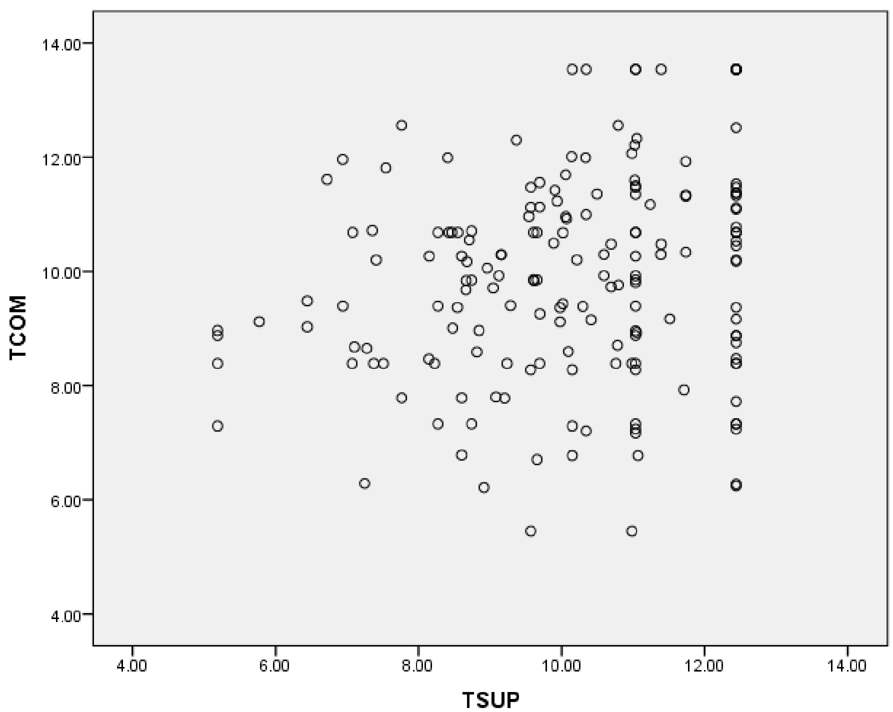


Figure C.13. Scatterplot of TCOM and TSUP (TIMSS 2015 Advanced)

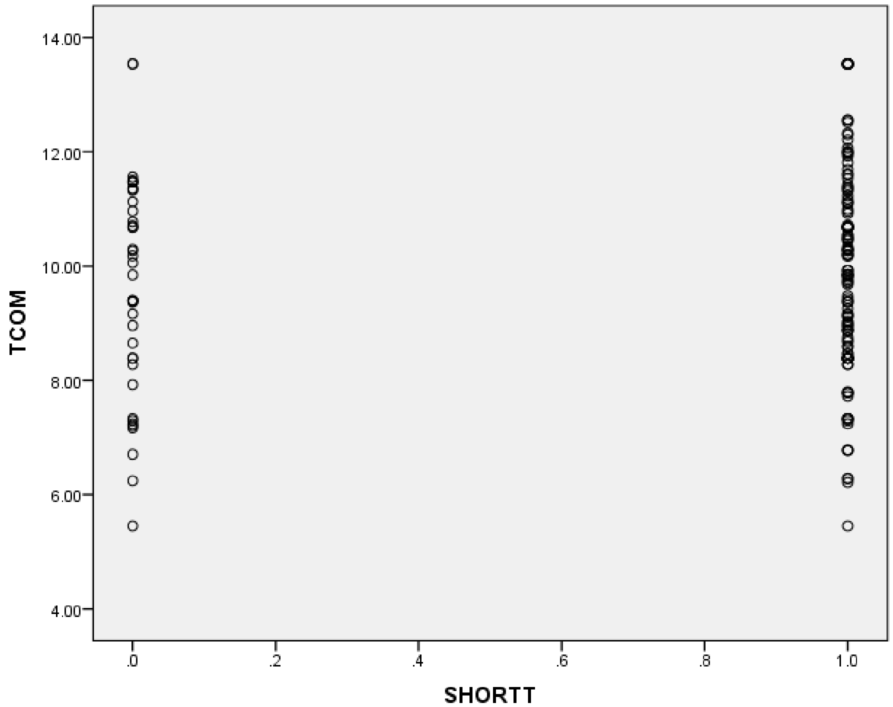


Figure C.14. Scatterplot of TCOM and SHORTT (TIMSS 2015 Advanced)

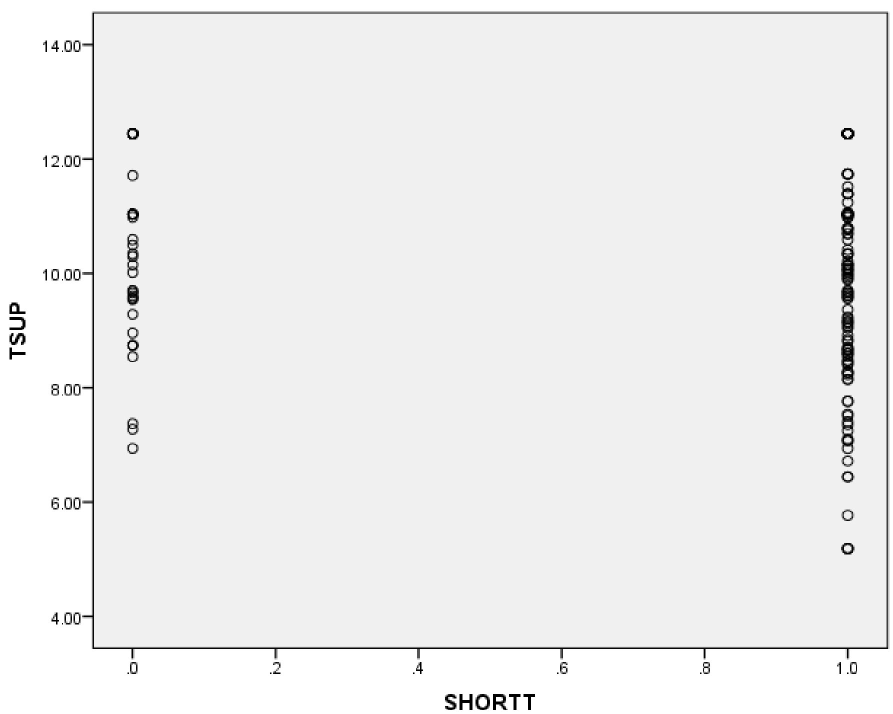


Figure C.15. Scatterplot of TSUP and SHORTT (TIMSS 2015 Advanced)

References

Aiken, L. R. (1972). Research on attitudes toward mathematics. *Arithmetic Teacher, 19*, 229 – 234.

Ames, C, & Archer, J. (1988). Achievement goals in the classroom: Student learning strategies and motivation processes. *Journal of Educational Psychology, 80*, 260-267

Ashcraft, M. H., & Kirk, E. P. (2001). The relationship among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General, 130*, 224 –237.

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review, 64*, 359-372.

Baines, E., Blatchford, P., & Kutnick, P. (2003). Changes in grouping practices over primary and secondary school. *International Journal of Educational Research, 39,* 9-34.

Baines, E., Blatchford, P., & Kutnick, P. (2008). Pupil grouping for learning: developing a social pedagogy in the classroom. In R. Gillies, A. Ashman, & J. Terwel (Eds.), *The teacher's role in implementing cooperative learning in the classroom* (pp. 55e71). New York: Springer.

Baker, D. P., & Deborah P. J. (1993). Creating Gender Equality: Cross-National Gender Stratification and Mathematical Performance. *Sociology of Education, 66,* 91–103.

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice Hall.

Baron-Cohen, S. (2003). The essential difference: The truth about the male and female brain. New York: Basic Books.

Battistich, V., Solomon, D., Kim, D., Watson, M., & Schaps, E. (1995). Schools as communities, poverty levels of student populations, and students' attitudes, motives, and performance. *American Educational Research Journal, 32,* 627-658.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 1860–1863. doi:10.1073/pnas.0910967107

Beller, M., & Gafni, N. (2000). Can item format (multiple choice vs open-ended) account for gender differences in mathematics achievement? *Sex Roles: A Journal of Research, 42(1-2),* 1-21.

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science, 210,* 1262–1264.

Bessudnov, A., & Makarov, A. (2013). Gender differences in mathematical performance and the school context: Evidence from Russia. *Higher School of Economics Research,* Retrieve from http://dx.doi.org/10.2139/ssrn.2265376

Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science, 355,* 389–391.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In Lord, F. M. and Novick, M. R. (eds.) *Statistical Theories of Mental Test Scores*, pp 397–479. Reading,MA: Addison-Wesley

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443-459.

Bong, M., and Clark, R. E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychology, 3,* 139–154.

Botella, C., Rueda, S., López-Iñesta, E., & Marzal, P. (2019). Gender Diversity in STEM Disciplines: A Multiple Factor Problem. *Entropy, 21,* 30.

Bryk, A. S., & Raudenbush, S. W. (1992). Hierarchical Linear Models: Applications and data analysis methods. Newbury Park, CA: Sage Publications.

Burton, N. W., and Lewis, C. (1996). Gender differences in college mathematics grades and SAT-M scores: A renanlysis of Wainer and Steinberg. *Journal of Educational Measurement, 33,* 257-270.

Butler, D. (2016). Tomorrow's world: technological change is accelerating today at an unprecedented speed and could create a world we can barely begin to imagine. *Nature, 530*, 398-402.

Carlton, S. T., & Harris, A. M. (1989, March). Female/male performance differences on the SAT: Causes and correlates. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Casey, M. B., Nuttall, R. L., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: A comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology, 33,* 669–680.

Campbell, J. R., & Beaudry, J. S. (1998) Gender Gap Linked to Differential Socialization for High-Achieving Senior Mathematics Students, *The Journal of Educational Research, 91*, 140-147, DOI: 10.1080/00220679809597534

Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software, 48(6),* 1-29. <doi:10.18637/jss.v048.i06>

Chipman, S. F., Krantz, D. H., & Silver, R. (1992). Mathematics anxiety and science careers among able college women. *Psychological Science, 3*, 292–295.

Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among 15-year-olds in 34 countries. *Learning and Instruction, 20,* 2–17.

Chudgar, A., & Sankar, V. (2008). The relationship between teacher gender and student achievement: evidence from five Indian states. *Compare: A Journal of Comparative and International Education, 38(5),* 627-642.

Deary, I. J., Thorpe, G., Wilson, V., Starr, J. M. & Whalley, L. J. (2003). Population sex differences in IQ at age 11: The Scottish mental survey 1932. *Intelligence, 31*, 533-542.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review, 95(2),* 158–165.

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *The Journal of Human Resources, 42(3),* 528–554.

Dee, T.S. (2006). The why chromosome: How a teacher's gender affects boys and girls. *Education Next, 6(4),* 68.

Devine, A., Fawcett, L., Szucs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and Brain Functions, 8,* 33–41.

Dweck, C. S., & Leggett, E. L. (1988). A social–cognitive approach to motivation and personality. *Psychological Review, 95,* 256–273.

Eccles, J. S. (1983). Expectancies, values, and academic behavior. In J. T. Spencer (Ed.), *Achievement and achievement motivation* (pp. 75-146). San Francisco: W. H. Freeman.

Eccles, J. S., and Jacobs, J. E. (1986). Social forces shape math attitudes and performance. *Signs, 11*, 367-380.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., et al. (1993). Development during adolescence: The impact of stage–environment fit on young adolescents' experiences in schools and in families. *American Psychologist, 48*, 90–101.

Eccles, J., Janis, E., & Harold, R. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues, 46*, 183–201.

Ellison, G., & Swanson, A. (2010). The gender gap in secondary school mathematics at high achievement levels: evidence from the American Mathematics Competitions. *Journal of Economic Perspectives, 24,* 109-128.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*, 101–127.

Entwisle, D. R., & Baker, D. P. (1983). Gender and young children's expectations for performance in arithmetic. *Developmental Psychology, 19,* 200–209.

Fassinger, P. A. (l995b). Understanding classroom interaction. Journal of Higher Education, 66, 82-96.

Fast, L. A., Lewis, J. L., Bryant, M. J., Bocian, K. A., Cardullo, R. A., Rettig, M., & Hammond, K. A. (2010). Does math self-efficacy mediate the effect of the perceived classroom environment on standardized math test performance? *Journal of Educational Psychology, 102,* 729-740.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin, 116,* 429–456.

Ferguson, R.F. (1991). Racial Patterns in How School and Teacher Quality Affect Achievement and Earnings. *Challenge, 2,* 1-35.

Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis, 21*, 97–110.

Foy, P., & Yin, L. (2016). Scaling the TIMSS 2015 Achievement Data. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 13.1-13.62). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timss.bc.edu/publications/timss/2015-methods/chapter-13.html

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through dolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology, 38*, 519–533.

Frost, L.A., Hyde, J.S., & Fennema, E. (1994). Gender, mathematics performance, and mathematics related attitudes and affect: a meta-analytic synthesis, *International Journal of Educational Research 21*, 373-385.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Allinder, R. M. (1991). The contribution of skills analysis within curriculum-based measurement in spelling. *Exceptional Children, 57,* 443– 452.

Gabriele, A. J., & Joram, E. (2007). Teachers' reflections on their reform-based teaching in mathematics: Implications for the development of teacher self-efficacy. *Action in Teacher Education, 29,* 60–74.

Gallagher, A. (1998). Gender and antecedents of performance in mathematics testing. *Teachers College Record, 100,* 297–314.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology, 75*, 165-190.

Gallagher, A., & DeLisi, R. (1994). Gender differences in Scholastic Aptitude Test-Mathematics problem solving among high-ability students. *Journal of Educational Psychology, 86,* 204–211.

Geary, D. C. (1998). Male, female: The evolution of human sex differences. Washington, DC: American Psychological Association.

Gentry, M., & Owen, S. V. (2004). Secondary student perceptions of classroom quality: Instrumentation and differences between advanced/honors and nonhonors classes. *The Journal of Secondary Gifted Education*, *16*, 20–29.

Goddard, R. D., Hoy, W. K., & Woolfolk Hoy, A. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal, 37,* 479–507.

Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers College Record, 109*, 877-896.

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do girls really experience more anxiety in mathematics? Psychological Science, 24, 2079–2087.

Good, T. L. (1981). Teacher expectations and student perceptions: A decade of research. *Educational Leadership, 55,* 415-423.

Goodenow, C. (1993). Classroom belonging among early adolescent students: Relationships to motivation and achievement. *Journal of Early Adolescence, 13*, 21-43.

Grobel, J., & Schwarzer, R. (1982). Social comparison, expectations and emotional reactions in the classroom. *School Psychology International, 3,* 49-56.

Guay, F., Marsh, H.W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology, 95*, 124-136.

Guskey, T. R. (1988). Teacher efficacy, self-concept, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education, 4,* 63–69.

Hackett, G. (1985). The role of mathematics self-efficacy in the choice of math-related majors of college women and men: A path analysis. *Journal of Counseling Psychology, 32,* 47– 56.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19,* 141-164.

Hargreaves, M., Homer, M., & Swinnerton, B. (2008). A comparison of performance and attitudes in mathematics amongst the 'gifted'. Are boys better at mathematics or do they just think they are? *Assessment in Education: Principles, Policy & Practice, 15,* 19–38.

Haynes, A. F., Anita, G. M., & Barry, S. S. (2004). Differential models for math anxiety in male and female college students, *Sociological Spectrum*, *24*, 295-318.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269,* 41-45.

Hembree, R. (1990). The nature, effects and relief of mathematics anxiety. *Journal for Research in Mathematics Education, 21*, 33-46.

Hoge, D. R., Smit, E. K., & Hanson, S. L. (1990). School experiences predicting changes in self-esteem of sixth and seventh-grade students. *Journal of Educational Psychology, 82,* 117–127. doi:10.1037//0022-0663.82.1.117

Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences, USA, 106,* 8801– 8807

Hyde, J. S., Fennema, E. H., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107,* 139–155.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A., & Williams, C. (2008). Gender similarities characterize math performance. *Science, 321,* 494–495.

Ing, M., Webb, N. M., Franke, M. L., Turrou, A. C., Wong, J., Shin, N., & Fernandez, C. H. (2015). Student participation in elementary mathematics classrooms: The missing link between teacher practices and student achievement? *Educational Studies in Mathematics, 90(3),* 341–356. doi:10.1007/s10649-015-9625-z

Johnson, B., & Stevens, J. J. (2006). Student achievement and elementary teachers' perceptions of school climate. *Learning Environments Research, 9,* 111–122. doi:10.1007/s10984-006-9007-7

Evetts, J. (1993) Women in Engineering: educational concomitants of a non-traditional career choice, *Gender and Education, 5*, 167-178, DOI: 10.1080/0954025930050204

Junge, M. E., & Dretzke, B. J. (1995). Mathematical self-efficacy gender differences in gifted/talented adolescents. *Gifted Child Quarterly, 39,* 22-2

Kimball, M. M. (1989). A new perspective on women's math achievement. *Psychological Bulletin, 105,* 198 –214.

Kotovsky, L. & Baillargeon, R. (1998). The development of calibration-based reasoning about collision events in young infants. *Cognition, 67*, 311-351.

Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis, 36(4), 476*–500.

Krieg, J. M. (2005). Student gender and teacher gender: what is the impact on high stakes test scores? *Current Issues in Education* [On-Line], 8(9). http://cie.ed.asu.edu/volume8/number9/.

Lam, Y. H. R., Tse, S. K., Lam, J. W. I., & Loh, E. K. Y. (2010). Does the gender of the teacher matter in the teaching of reading literacy? Teacher gender and pupil attainment in reading literacy in Hong Kong. *Teaching and Teacher Education, 26*, 754-759

Laukaityte, I., & Wiberg, M. (2017). The Importance of Sampling Weights in Multilevel Modeling of International Large-Scale Assessment Data. *Communications in Statistics - Theory and Methods, 47,* 4991-5012, https://doi.org/10.1080/03610926.2017.1383429

Learmonth, A. E., Newcombe, N. S., & Huttenlocher, J. (2001). Toddlers' use of metric information and landmarks to reorient. *Journal of Experimental Child Psychology, 80,* 225-244.

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences, 19,* 355-365.

Lee, V. E., & Smith, J. B. (1997). High school size: Which works best for whom? *Educational Evaluation and Policy Analysis, 19,* 205–227.

Legewie, J., & DiPrete, T. A. (2014). The high school environment and the gender gap in science and engineering. *Sociology of Education, 87*, 259–280.

Lent, R. W., Brown, S. D., & Larkin, K. C. (1986). Self-efficacy in the prediction of academic performance and perceived career options. *Journal of Counseling Psychology, 33,* 265–269.

Lindberg, S. M., Hyde, J. S., & Petersen, J. L. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*, 1123–1135.

Lindberg, S., Linkersdorfer, J., Ehm, J.-H., Hasselhorn, M., & Lonnemann, J. (2013). Gender differences in children's math self-concept in the first years of elementary school. *Journal of Education and Learning, 2,* 1-8.

Liu, O. L., & Wilson, M. (2009a). Gender differences in large-scale mathematics assessments: PISA trend 2000 & 2003. *Applied Measurement in Education, 22,* 164–184.

Lu¨dtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology, 34,* 120–131. doi:10.1016/j.cedpsych.2008.12.001

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education, 30,* 520–540.

Ma, X. (2008). Within-school gender gaps in reading, mathematics, and science literacy. *Comparative Education Review, 52*, 437–460.

Ma, X., & Willms, J. D. (2004). School disciplinary climate: Characteristics and effects on eighth grade achievement. *Alberta Journal of Educational Research, 50*, 169-188.

Maccoby, E. E., & Jacklin, C. H. (1974). The psychology of sex differences. Stanford, CA: Stanford University Press

Mark, G. N. (2008). Accounting for the gender gaps in student performance in reading and mathematics: Evidence from 31 countries. *Oxford Review Education, 34*, 89-109.

Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and English constructs. *American Educational Research Journal, 35,* 705-738.

Martin, M. O., Mullis, I. V. S., Hooper, M., Yin, L., Foy, P., & Palazzo, L. (2016). Creating and Interpreting the TIMSS 2015 Context Questionnaire Scales. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 15.1-15.312). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timss.bc.edu/publications/timss/2015-methods/chapter-15.html

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149-174.

McGraw, R., Lubienski, S. T., & Strutchens, M. E. (2006). A closer look at gender in NAEP mathematics achievement and affect data: Intersections with achievement, race and socio-economic status. *Journal for Research in Mathematics Education, 37,* 129–150.

Meece, J. L., Wigfield, A., Eccles, J. S. (1990). Predictors of math anxiety and its consequences for young adolescents' course enrollment intentions and performances in mathematics. *Journal of Educational Psychology*, *82*, 60–70.

Meece, J., Blumenfeld, P., & Hoyle, R. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology, 80,* 514-523

Michaelowa, K. (2001). Primary education quality in francophone sub-Saharan Africa: determinants of learning achievement and efficiency considerations. *World Development, 29(10),* 1699-1716

Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). . Change in teacher efficacy and student self-and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology, 81,* 247-258.

Mills, C. J., Ablard, K. E., & Stumpf, H. (1993). Gender differences in academically talented young students' mathematical reasoning: Patterns across age and subskills. *Journal of Educational Psychology, 85,* 340–346.

Monk, D. H., & Haller, E. J. (1993). Predictors of high school academic course offerings: The role of school size. *American Educational Research Journal, 30,* 3-21.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children, 5,* 113-127.

Murdock, T. B., & Miller, A. (2003). Teachers as sources of middle school students' motivational identity: Variable-centered and person-centered analytic approaches. *Elementary School Journal, 103, 3*83–399.

National Science Foundation. (2016). Science and Engineering Indicators 2016. Arlington, VA: Author.

O'Neill, K. A., & McPeek, W. M. (1993). Item and test characteristics that are associated with differential item functioning. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 255–279). Hillsdale, NJ: Erlbaum.

O'Neill, K. A., Wild, C. L., & McPeek, W. M. (1989). Gender-related differential item performance on graduate admissions tests. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology, 86,* 193–203.

Pajares, F., & Miller, M. D. (1997). Mathematics self-efficacy and mathematical problem solving: Implications of using different forms of assessment. *Journal of Experimental Education, 65,* 313-229.

Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology, 20,* 426-443.

Parsons, J. E., Adler, T. F., & Kaczala, C. M. (1982). Socialization of achievement attitudes and beliefs: Parental influences. *Child Development, 53,* 310-321.

Parsons, Jacquelynn Eccles, Kaczala, Caroline M., & Meece, Judith L. (1982). Socialization of achievement attitudes and beliefs: Classroom influences. *Child Development, 53,* 322–339

Patrick, H., Ryan, A. M., & Kaplan, A. (2007). Early adolescents' perceptions of the classroom social environment, motivational beliefs, and engagement. *Journal of Educational Psychology, 99,* 83-98.

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal, 45, 365*-397.

Pomerantz, E. M., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance and internal distress. *Journal of Educational Psychology, 94,* 396–404.

Potter, F. (1988). Survey of Procedures to Control Extreme Sampling Weights, ASA Proceedings of the Section on Survey Research Methods, American Statistical Association, 1988, 453-458.

Potter, F. (1990). A Study of Procedures to Identify and Trim Extreme Sampling Weights. Proceedings of the Section on Survey Research Methods, American Statistical Association, 1990, 225-230.

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory and math achievement in early elementary school. *Journal of Cognition and Development, 14,* 187–202.

Randhawa, B. S. (1994). Self-efficacy in mathematics, attitudes, and achievements of boys and girls from restricted samples in 2 countries. *Perceptual and Motor Skills, 79,* 1011–1018.

Randhawa, B. S., Beamer, J. E., & Lundberg, I. (1993). Role of mathematics self-efficacy in the structural model of mathematics achievement. *Journal of Educational Psychology, 85,* 41–48.

Rasch, G. (1960/1980). Probabilistic models for some intelligence and attainment tests. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.

Raudenbush, S. W., Bryk, A., & Congdon, R. (2005). HLM Statistical Software: Version 6. Lincolnwood, IL: Scientific Software International, Inc. Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage.

Reddy, R., Rhodes, J., & Mulhall, P. (2003). The influence of teacher support on student adjustment in the middle school years: A latent growth curve study. *Development and Psychopathology, 15, 119–138.*

Reis, S. M., & Park, S. (2001). Gender differences in high-achieving students in math and science. *Journal for the Education of the Gifted, 25,* 52–73.

Richardson, F.C., & Suinn, R.M. (1972). The Mathematics Anxiety Rating Scale: Psychometric data. *Journal of Counseling Psychology, 19*, 551 – 554.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (1998). Teachers, schools, and academic achievement. (Working Paper No. 6691). Cambridge, MA: National Bureau of Economic Research.

Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal, 52(3),* 475–514.

Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the Prospects Study of Elementary Schools. *Teachers College Record, 104,* 1525–1567.

Ryan, A. M., & Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal, 38,* 437–460.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph Supplement, 17* (4, Pt. 2).

Samejima, F. (1996). Graded response model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85-100). New York: Springer.

Schunk, D. H., & Lilly, M. W. (1984). Sex differences in self-efficacy and attributions: Influence of performance feedback. *Journal of Early Adolescence, 4,* 203-213.

Schunk, D. H., and Ertmer, P. (2000). Self-regulation and academic learning: Self-efficacy enhancing interventions. In Boekaerts, M., Pintrich, P. R., and Zeidner, M. (eds.), *Handbook of Self-Regulation*, Academic Press, San Diego, CA, pp. 631–649.

Seegers, G., & Boekaerts, M. (1996). Gender-related differences in self-referenced cognitions in relation to mathematics. *Journal for Research in Mathematical Education, 27,* 215-240.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10,* 80-83.

Siegel, R. G., Galassi, J. P., & Ware, W. B. (1985). A comparison of two models for predicting mathematics performance: Social learning versus math aptitude-anxiety. *Journal of Counseling Psychology, 32,* 531–538.

Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles: A Journal of Research, 50,* 241-252.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85,* 571 – 581.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance.

Journal of Experimental Social Psychology, 35, 4-28.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613-629.

Stipek, D. J. (2002). Motivation to learn (4th ed). Boston: Allyn & Bacon.

Swann, J. & Graddol, D. (1988) Gender inequalities in classroom talk, *English in Education, 22,* 48–65.

Tiedermann J. (2000) Parent's gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology, 92,* 144–151.

TIMSS & PIRLS International Study Center (2016). TIMSS 2015 Achievement Scaling Methodology. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 12.1-12.9). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timss.bc.edu/publications/timss/2015-methods/chapter-12.html

TIMSS & PIRLS International Study Center (2016). TIMSS 2015 Achievement Scaling Methodology. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 12.1-12.9). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timss.bc.edu/publications/timss/2015-methods/chapter-12.html

Trust. (n.d.). In Merriam-Webster's online dictionary (11th ed.). Retrieved from http://www.m-w.com/dictionary/trust

Turner, S.L., Steward, J.C., & Lepan, R.T. (2004). Family factors associated with sixth-grade adolescents' math and science career interests. *The Career Development Quarterly, 53,* 41–52.

Van Houtte, M. (2007). Exploring teacher trust in technical/vocational secondary schools: Male teachers' preference for girls. *Teaching and Teacher Education, 23,* 826-839.

Webb, N. M., Franke, M. L., Ing, M., Turrou, A. C., Johnson, N. C., & Zimmerman, J. (2019). Teacher practices that promote productive dialogue and learning in mathematics classrooms. *International Journal of Educational Research, 97,* 176-186.

Whitbeck, L. B., & Gecas, V. (1988). Value attributions and value transmission between parents and children. *Journal of Marriage and the Family, 50*, 829–840.

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review, 6*, 49–78.

Wigfield, A., & Eccles, J. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review, 12*, 265–310.

Wigfield, A., & Meece, J. (1988). Math anxiety in elementary and secondary school students. *Journal of Educational Psychology, 80,* 210–216.

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods, 12,* 58–79. doi:10.1037/1082-989X.12.1.

Yee, D. K., & Eccles, J. S. (1988). Parent perceptions and attributions for children's math achievement. *Sex Roles*, *19*, 317-333.

Younger, M., Warrington, M., & Williams, J. (1999). The gender gap and classroom interactions: reality and rhetoric? *British Journal of Sociology of Education*, *20(3),* 325–341.

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology, 81,* 329-339.