

Comparing Male and Female Student Self-Efficacy, Identification,
And Achievement In Mathematics: A Cross-Lagged Panel Analysis
of Causal Effects

By
© 2020

Cassandra Mia Bonitto
Ed.S., University of Kansas, 2014
B.A., Wichita State University, 2011

Submitted to the graduate degree program in School Psychology and the Graduate Faculty of
the University of Kansas in partial fulfillment of the requirements for the degree of Doctor
of Philosophy.

Chair: Matthew R. Reynolds, Ph.D.

Christopher R. Niileksela, Ph.D.

Meagan M. Patterson, Ph.D.

Kristen N. Hensley, Ph.D.

Lisa E. Wolf-Wendel, Ph.D.

Date Defended: 1 June 2020

The dissertation committee for Cassandra Mia Bonitto certifies that this is
the approved version of the following dissertation:

**Comparing Male and Female Student Self-Efficacy, Identification,
And Achievement In Mathematics: A Cross-Lagged Panel Analysis
of Causal Effects**

Chair: Matthew Reynolds, Ph.D.

Date Approved: 8 June 2020

Abstract

Understanding the direction of causality in the relationship between self-beliefs and academic performance is essential in educational research. The two purposes of this study were to examine the temporal relationships between mathematical identity, mathematical self-efficacy, and mathematics achievement and to understand whether those relationships were moderated by student gender. A cross-lagged panel analysis of causal effects was conducted using data from the High School Longitudinal Study of 2009 to examine the key variables in 9th and 11th graders. Then models of female and male students were developed, and tests of moderation were conducted. The cross-lagged panel analysis found support for the skill-development model of self-efficacy (i.e., higher achievement results in increases in self-efficacy) and found a reciprocal relationship between mathematical identity and mathematics achievement (i.e., higher identity results in increases in achievement and higher achievement result in increases in identity). Also, a one-way relationship between mathematical identity and mathematical self-efficacy was found, showing that 9th grade mathematical identity influences 11th grade mathematical self-efficacy. The tests of moderation showed that gender moderated the relationship between 9th grade mathematics achievement and 11th grade mathematical self-efficacy and 11th grade mathematical identity. In addition, mathematical identity was found to be more stable for 9th grade female students. These findings show that female student's development of mathematical identity and self-efficacy were less influenced by mathematics achievement success than they are in male students. The data gathered from this study will inform researchers, policymakers, and educators about the paths that male and female high school students' follow to succeed in mathematics.

Acknowledgments

It would be impossible to communicate all the people who have contributed to my education. The completion of a doctorate is not a singular accomplishment but is rather done through small, everyday efforts from many people.

First, I would like to thank Dr. Reynolds, my dissertation advisor. When I first applied to KU, I barely understood what I was asking for when I requested for him to be my advisor. Since then, he has patiently guided me, despite my resistance, through the process of becoming a researcher and psychologist. I use the information and lessons you taught me every day. I also want to thank the rest of my dissertation committee. Thank you to Dr. Patterson, who supervised me as a graduate teaching assistant, just out of my undergraduate studies, and taught me to be a competent teacher. Thank you to Dr. Niileksela, who allowed me to supervise his students in his clinic class, providing me with my first experience of supervising school psychology students. Thank you to Dr. Hensley, who supervised my clinic experience in my second year in the program. Finally, thank you to Dr. Wolf-Wendel for all your thoughtful feedback on my dissertation.

Next, I would like to thank the faculty outside of KU, who contributed to this degree. Specifically, the faculty at my undergraduate institution, Wichita State University. Especially Dr. Jeffrey Jarman, for being a mentor throughout my experiences at WSU. I also want to thank Dr. Brenda Huber, my doctoral internship supervisor at Illinois State University. Thank you both for your ongoing guidance and friendship.

Next, I would like to thank my friends. First, I'd like to recognize my entire Illinois School Psychology Internship Consortium (ISPIC) cohort for cheering me on through this. I'd especially like to thank Dr. Paula Allee-Smith, Dr. Jeffrey Garofano, Dr. Daniel Philippe, Dr.

Kristy Warmbold, Dr. Erin Yosai, Dr. Leah Marks, Dr. Jared Bishop, and Dr. Geovanna Rodriguez for their ongoing friendship. I appreciate getting to be a part of the ISPIC network. Thank you to Dr. Meghan Ecker for your support and guidance in completing the statistics in this study. Thank you to Rachel Bromberg for the emotional support, friendship, and jokes that kept me laughing through this journey. Thank you to Em Meyer for your friendship and support throughout the dissertation process. Thank you to Monika Jasso and Dr. Angela Beeler. I could not have finished my dissertation while working as a professor without the support of you two as coworkers in school psychology. Finally, thank you to Dustin Wiens, who desperately wanted to help me title this dissertation, and spent many nights suggesting math puns. While none of them ended up as the title, I did appreciate your additions.

Lastly, I need to thank my parents and family for their love and support. I especially want to recognize my sister, Ceceli Bonitto, who will soon be finishing her Doctorate of Audiology. The Drs. Bonitto will be a force, and I am so excited for you! Most of all, I need to thank Matthew Coleman. You are an incredible life partner. You kept me grounded and supported me through this process. There is not enough space to list everything you've done. Thank you. I love you.

Dedication

This dissertation is dedicated to the women, transgender men, and nonbinary folks who identify with activities thought of as belonging to cisgender men. Whether you are on the debate team, on the football field, or in a STEM field, I see you, I believe in you, and I hope you will continue to thrive while doing what you love.

Table of Contents

Abstract	iii
Acknowledgments.....	iv
Dedication	vi
Table of Contents	vii
List of Figures	xi
List of Tables	xii
Chapter 1: Introduction	1
Purpose of the Study	1
Mathematical Self-Efficacy and Mathematics Achievement	1
Mathematical Identity and Mathematics Achievement	3
Mathematical Identity and Mathematical Self-Efficacy	3
Mathematical Identity, Mathematical Self-Efficacy, and Mathematics Achievement.....	4
Gender’s Relationships with Mathematical Identity, Mathematical Self-Efficacy, and Mathematics Achievement	5
Problem Statement	7
Research Questions	9
Chapter 2: Literature Review.....	10
Self-Efficacy, Self-Esteem, and Self-Concept.....	10
Self-Efficacy and Achievement	12
Relation between self-efficacy and achievement.....	12
Self-efficacy and achievement theory.....	13
Growth in academic self-efficacy in school-age students.....	15

Co-development of self-efficacy and achievement.....	16
Mathematical Self-Efficacy and Mathematics Achievement	18
Mathematical self-efficacy and mathematics achievement longitudinal research.....	19
Self-enhancement model.....	21
Reciprocal model	22
Mathematical Identity	24
Mathematical Identity and Mathematics Achievement	26
Mathematical Self-Efficacy and Mathematical Identity	30
Gender’s Relationship with Mathematical Self-Efficacy, Mathematical Identity & Mathematics Achievement	32
Chapter 3 Methods	38
Participants.....	38
Measures	39
Gender.....	39
Mathematics Achievement.....	40
Mathematical Self-Efficacy	42
Mathematical Identity	42
Missing Data	43
Analytic Procedure	44
Measurement Model	44
Factorial Invariance.....	45
Longitudinal Factorial Invariance	46
Multiple Group Factorial Invariance.....	47

Structural Model	47
Model Evaluation	49
Research Questions.....	50
Chapter 4: Results.....	52
Item Level Descriptive Statistics	52
Missing Data.....	55
Measurement Model	56
Research Question 1	59
Research Question 2	62
Measurement Models by Gender	62
Multi-Group Invariance Testing	66
Structural Model	67
Chapter 5: Discussion	71
Summary of Findings.....	71
Preliminary Analyses	71
Hypothesis 1.....	72
Hypothesis 2.....	73
Implications for Theory	74
Mathematical Self-Efficacy and Mathematics Achievement.....	74
Mathematical Identity and Mathematics Achievement	75
Mathematical Identity and Mathematical Self-Efficacy	77
Limitations	78
Implications for Future Research and Practice	80

Conclusions..... 82

References..... 83

List of Figures

Figure 1: <i>Latent variable cross-lagged panel model of the relationship of mathematical self-efficacy, mathematical identity, and mathematics achievement</i>	49
Figure 2. <i>Longitudinal Measurement Model for the Total Sample</i>	57
Figure 3. <i>Correlated Residual Longitudinal Measurement Model</i>	58
Figure 4. <i>Total Sample Longitudinal Structural Model</i>	60
Figure 5. <i>Longitudinal Measurement Model for the Female Student Sample</i>	63
Figure 6: <i>Longitudinal Measurement Model for the Male Student Sample</i>	63
Figure 7. <i>Female Correlated Residual Longitudinal Measurement Model</i>	65
Figure 8: <i>Male Correlated Residual Longitudinal Measurement Model</i>	65
Figure 9. <i>Female Student Structural Model</i>	68
Figure 10. <i>Male Student Structural Model</i>	69

List of Tables

Table 1. <i>Demographic Information</i>	39
Table 2. <i>Bivariate Correlations, Skewness & Kurtosis for Observed Variables</i>	53
Table 3. <i>Sample Size, Means, & Standard Deviations for Observed Variables in Total Sample and Male and Female Students</i>	54
Table 4. <i>Model Fit Statistics for Longitudinal Measurement Models with the Total Sample</i>	59
Table 5. <i>Unstandardized and Standardized Regression Weight Estimates of Temporal Relations Within and Between Constructs</i>	61
Table 6. <i>Model Fit Statistics for Male and Female Longitudinal Measurement and Structural Models</i>	62
Table 7. <i>Model Fit Statistics for Configural and Factorial Invariance and for Tests of Moderation Across Gender</i>	66
Table 8. <i>Unstandardized and Standardized Regression Weight Estimates by Gender</i>	67

Chapter 1: Introduction

Purpose of the Study

The question of the direction of causality in the relationship between self-beliefs and academic performance has been described as one of the “thorniest issues” in education research (Pajares & Schunk, 2001, p. 239). Two important self-belief constructs that have recently gained the interest of policymakers and educators are mathematical identity and mathematical self-efficacy (Trujillo & Tanner, 2014). The two purposes of this study are to investigate the longitudinal relationships between mathematical self-efficacy, mathematical identity, and mathematics achievement, with an emphasis on understanding how these effects operate in high school students, and whether the longitudinal effects work in the same ways for male students as they do for female students. The results of this study will allow policymakers and educators to understand better the path students follow to succeed in mathematics.

Mathematical Self-Efficacy and Mathematics Achievement

Mathematical self-efficacy is an internal assessment of competence in mathematics (Möller, Pohlmann, Köller, & Marsh, 2009). Empirical research conclusively demonstrates that there is a positive relationship between mathematical self-efficacy and mathematics achievement (Möller et al., 2009). Three theories have been developed to explain the nature of the relationship. The first theory is the self-enhancement model. In this theory, mathematical self-efficacy affects later mathematics achievement: If a student believes that they can be successful in mathematics, they are more likely to engage in challenging mathematics tasks, and thus develop their mathematics skills (Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005). The second theory is the skill development model. In this theory, mathematics achievement affects mathematical self-efficacy: If a student experiences success in mathematics tasks, then they are

more likely to predict themselves to be successful in mathematics in the future (Ganley & Lubienski, 2016; Skaalvik & Valås, 1999). The third theory is the reciprocal model. In this theory, mathematics achievement and mathematical self-efficacy have a mutually reinforcing relationship (Huang, 2011; Pinxten, Marsh, De Fraine, Noortgate, & Van Damme, 2014; Valentine, DuBois, & Cooper, 2004): Mathematical self-efficacy encourages persistence in mathematics tasks, while persistence in the tasks increases the probability of achievement, which in turn increases the likelihood that the student would anticipate future mathematics success (Pinxten et al., 2014). Given the empirical support for all of the directional models, the true direction of the relationship between mathematical self-efficacy and mathematics achievement is uncertain. Despite inconsistent findings, however, there seems to be the most support for the reciprocal model, both empirically and theoretically (Huang, 2011; Valentine et al., 2004).

Although there is a large amount of research examining self-efficacy as it relates to mathematics achievement, many of the relevant longitudinal studies have been conducted with international populations (Marsh et al., 2005; Pinxten et al., 2014; Skaalvik & Valås, 1999), students in elementary and middle school (Ganley & Lubienski, 2016; Marsh et al., 2005; Pinxten et al., 2014; Pomerantz, Altermatt, & Saxon, 2002), and in college-age populations (Marshman et al., 2018; Robnett, Chemers, & Zurbriggen, 2015). Few studies have examined the longitudinal relations between these variables with high school students in the United States. High school is an essential time for the development of self-beliefs (Klimstra, Hale, Raaijmakers, Branje, & Meeus, 2010); therefore, educators and policymakers need to understand better the nature of these relationships with this specific population.

Mathematical Identity and Mathematics Achievement

There have been few empirical studies conducted on the construct of mathematical identity. Mathematical identity refers to a person's beliefs, attitudes, emotions, and dispositions about mathematics. Most simply, a person with a robust mathematical identity is likely to think of themselves as a "math person" (Froschl & Sprung, 2016, p. 321). Almost everything that has been written about mathematical identity is theoretical (Darragh, 2016; Radovic, Black, Williams & Salas, 2018). Studies of mathematical identity and mathematics achievement have been mostly qualitative (McGee & Martin, 2011; Radovic et al., 2018). Other than one study that found evidence of a small positive relationship between the two (Williams, Burt & Hilton, 2016), the empirical relationship between mathematical identity and mathematics achievement is unknown. Yet, theory suggests the possibility of a stronger relationship than reported in the Williams et al. (2016) study. When a student has high achievement in mathematics, they are more likely to identify with being a math person. High mathematics achievement may increase interest and also influence the person's self-perception. Likewise, when a student identifies as a math person, they are more likely to have high achievement in mathematics, as they are more likely to engage in mathematics activities, consistent with their identity (Grootenboer, Zevenbergen & Chinnappan, 2008). Thus, there is a reciprocal and theoretically possible mutually beneficial relationship between the two over time. Despite the theory, however, little is known about the size, direction, and nature of the longitudinal relationship between mathematical identity and mathematics achievement.

Mathematical Identity and Mathematical Self-Efficacy

Mathematical identity and mathematical self-efficacy likely have a positive relationship. Someone who experiences achievement success in mathematics is more likely to identify as a

math person, and someone who succeeds in mathematics likely believes they are competent in mathematics (Grootenboer, Zevenbergen & Chinnappan, 2008). From the few empirical studies of mathematical identity, mathematical identity, and mathematical self-efficacy do indeed appear to be strongly correlated. Huang, Cribbs, and Piatek-Jimenez (2016) found a large correlation ($r = .81$) between mathematical self-efficacy and mathematical identity in a sample of undergraduate students. In another study with high school students, after controlling for utility and interest in mathematics, mathematical self-efficacy was strongly related ($\beta = .58$) to mathematical identity (Middleton 2013). These findings suggest that mathematical self-efficacy and mathematical identity are strongly related but distinct concepts. The longitudinal relation between the two, however, is not known.

Mathematical Identity, Mathematical Self-Efficacy, and Mathematics Achievement

Although both mathematical self-efficacy and mathematical identity have been examined as they relate to each other, and as they relate to mathematics achievement, individually, few studies have explored all three variables in one model (Ganley & Lubienski, 2016; Middleton 2013; Middleton, Mangu, & Lee, 2019). The most related past study was conducted by Middleton (2013). This study examined mathematical identity, mathematical self-efficacy, and mathematics achievement of 9th grade students in a path model. It was not clear if mathematical identity was associated with mathematics achievement in the study, but it was related to mathematical self-efficacy. Correlations were not reported, however, and this study did not examine the relationships over time. Middleton, Mangu, and Lee (2019) completed another related study and did use longitudinal data. But in this study, they investigated how mathematical self-efficacy, mathematical identity, mathematical interest, and mathematical utility predicted science, technology, engineering, and math (STEM) career intentions and motivation. This

relationship between these variables and mathematics achievement was not studied. Thus, little is known about the relationship between mathematical identity and mathematical self-efficacy and the relationship between mathematical identity and mathematics achievement, let alone a possible longitudinal relationship.

To summarize, research has generally supported a reciprocal model that describes the longitudinal relationship between mathematical self-efficacy and mathematics achievement (Huang, 2011; Valentine et al., 2004). Because mathematical self-efficacy and mathematical identity are potentially highly related (Williams, Burt & Hilton, 2016; Middleton 2013) and likely have similar relationships with mathematics achievement, any study of the relationship between mathematical identity and mathematics achievement should demonstrate the relationship while also controlling for mathematical self-efficacy. That is, it is essential to understand whether the relationship between mathematical identity and mathematics achievement is unique and not merely a function of mathematical self-efficacy.

Gender's Relationships with Mathematical Identity, Mathematical Self-Efficacy, and Mathematics Achievement

Research has shown that female students have both lower mathematical self-efficacy (Brainard & Carlin, 1998; Else-Quest, Hyde & Linn, 2010) and lower identification with STEM fields than male students (Hazari, Sonnert, Sadler & Shanahan, 2010). These findings persist despite evidence that, when controlling for access to education, female students tend to achieve at approximately the same level as male students on most mathematics tests and grades (Else-Quest et al., 2010). Lower levels of mathematical self-efficacy for female students have been documented even in samples in which the female students obtain higher mathematics grades than male students (Pomerantz et al., 2002). Lower levels of mathematical self-efficacy and

mathematical identity are associated with a lower probability of engaging in STEM experiences, choosing STEM careers, and pursuing advanced STEM education (Middleton, Mangu, & Lee, 2019). This has produced a gender gap in representation in advanced STEM education and STEM career fields (Botella, Rueda, López-Iñesta, & Marzal, 2019).

One recent study conducted by Ganley and Lubienski (2016) examined gender differences in the longitudinal relationships between mathematical self-efficacy, mathematical interest, and mathematics achievement in 3rd to 8th grade students. Results indicated that gender differences in mathematical self-efficacy were more extensive than gender differences in interest and achievement in middle school. Although female students had lower self-efficacy than male students, the longitudinal relationships between the variables for male and female students did not differ. These findings are similar to other studies, which found mean differences in mathematical self-efficacy, yet no moderating impact of gender on the relationship between mathematical self-efficacy and mathematics achievement (Else-Quest et al., 2010; Marsh et al., 2005; Pomerantz et al., 2002).

Despite the lack of documentation of a moderating relationship, there is still more to examine here, with the potential for a significant relationship to be found. The vast majority of longitudinal studies in this area were conducted with college students (Marsh et al., 2005; Pomerantz et al., 2002; Brainard & Carlin, 1998) or examined international populations (Else-Quest et al., 2010). More information is needed on U.S. high school aged students. High school is a crucial age for the development of self-beliefs (Klimstra et al., 2010); therefore, insight into this population may help to guide U.S. policymakers and educators.

While there is more extensive research on mathematical self-efficacy and mathematics achievement, no longitudinal studies of mathematical identity and mathematics achievement

were found that examined gender differences. The only relevant studies found examined two large, national samples of college students, and found that men are about half a standard deviation more likely to perceive themselves as a physics person than women (Hazari et al., 2010; Hazari, Sadler & Sonnert, 2013). While this study suggests the possibility of similar mean differences to mathematical self-efficacy, the size, nature, and direction of the longitudinal relationship are unknown and worthy of study.

Finally, it is essential to note here that society's understanding of sex and gender has evolved since 2009 when the dataset used in this study was initially gathered. Given how the sex/gender variable was assessed in the dataset, students were forced to identify as male or female while completing the survey. This is undoubtedly a limitation of this study, as it does not allow conclusions to be made about students who are intersex, transgender, non-binary, or otherwise do not identify as male or female. Despite this, there is still value in studying students who identify as male or female, as this does provide some insights, even if limited, into differences that may be caused by the different social experiences of these groups of students (Hyde, Bigler, Joel, Tate, & Van Anders, 2019).

Problem Statement

While theoretically, there is likely a longitudinal relationship between mathematical self-efficacy, mathematical identity, and mathematics achievement, more research is needed to determine the nature and direction of this relationship (Huang, 2011; Radovic et al., 2018). In particular, little at all is known about mathematical identity and its relationship with mathematics achievement and mathematical self-efficacy, as the vast majority of research on mathematical identity has been theoretical (Darragh, 2016). Also, while popular theory identifies an achievement based "gender gap," it appears that there may not be a difference in the way that

gender affects the relationship between mathematical self-efficacy and mathematics achievement. There is also limited research on this relationship in high school populations. As high school is an especially important time for identity and self-efficacy development, more research in this population is especially relevant (Klimstra et al., 2010).

There is also a need for a longitudinal analysis of the variables of mathematical self-efficacy, mathematical identity, and mathematics achievement in order to understand the nature and direction of the relationships between the variables and to understand better the role that gender plays in their relationship. This study will fill these gaps in the research, allowing educators and policymakers to understand better the relationship between mathematical self-efficacy, mathematical identity, and mathematics achievement in U.S. high school students.

This dissertation will examine data from the High School Longitudinal Study of 2009 (HSLs:09), a nationally representative large scale survey of the variables of interest. The High School Longitudinal Study of 2009 (HSLs: 09) was conducted by the National Center for Education Statistics (NCES) in response to a policy mandate to collect and disseminate statistics and other data related to STEM education in the United States and the need for policy-relevant, nationally representative longitudinal samples of secondary school students (Ingels, Herget, Pratt, Copello, Leinwand, 2010). Researchers have been encouraged to use the High School Longitudinal Study of 2009 dataset to gain resolution on the variables involved in student's attitudes and decisions predicting achievement in STEM (Maltese & Tai, 2009). It also has the potential to better inform educators and policymakers about any barriers that may affect gender representation in collegiate STEM majors and STEM careers (Botella, Rueda, López-Iñesta, & Marzal, 2019).

Research Questions

Question 1. What is the temporal relationship between mathematical self-efficacy, mathematical identity, and mathematics achievement?

Question 2: Do the observed temporal relationships among mathematical self-efficacy, mathematical identity, and mathematics achievement depend on gender?

The next chapters will be as follows: Chapter 2 will explore the current research on mathematical self-efficacy, mathematical identity, and mathematics achievement, how the variables relate, and how gender may influence the relationship between the variables. Chapter 3 will explain the methods used in the study. Chapter 4 will describe the results of the analysis. Finally, Chapter 5 will summarize the findings, explore future research directions, and discuss the implications for theory and applications for educators and policymakers.

Chapter 2: Literature Review

The literature review will begin by defining and examining self-esteem, self-concept, self-efficacy, mathematical self-efficacy, and mathematics achievement. Then, there will be a discussion of how mathematical self-efficacy relates to mathematics achievement. Next, mathematical identity and its relationship to mathematics achievement and mathematical self-efficacy will be explored. Finally, the literature review will conclude by considering the way that gender may shape these relationships and influence the development of mathematical self-efficacy and mathematical identity in female students.

Self-Efficacy, Self-Esteem, and Self-Concept

Albert Bandura (1977) is credited with introducing the construct of self-efficacy. He was initially interested in studying self-esteem as it relates to human motivation, as he believed that people are motivated to act by their belief in themselves. However, as his research progressed, Bandura (1986) began to hypothesize that self-efficacy played a significant role in human motivation because “the types of outcomes people anticipate depend largely on their judgments of how well they will be able to perform in given situations” (p. 392). Since Bandura’s foundational research, the constructs of self-esteem and self-efficacy have often been confused because successful performance on a task can increase both self-efficacy and self-esteem (Zimmerman, 2000). Therefore, it is essential to distinguish between self-efficacy and self-esteem when examining these constructs. Self-efficacy is situationally specific, whereas self-esteem is a more broadly understood concept, referring to a highly favorable, global evaluation of the self (Baumeister, Campbell, Krueger, & Vohs, 2003). Baumeister (2005) defines self-esteem as “how people evaluate themselves” (p. 36). Trautwein, Lüdtke, Köller, and Baumert

(2006) used the term "self-esteem to describe the global perception of the self as a person" (p. 335). These definitions of self-esteem are, therefore, in contrast to the specificity of self-efficacy.

Definitions of self-efficacy have varied. Bandura (1997) defined self-efficacy as the belief in one's capabilities to organize and execute the courses of action required to produce given attainments. According to Pajares and Miller (1994), "self-efficacy is a context-specific assessment of competence to perform a specific task, a judgment of one's capabilities to execute specific behaviors in specific situations" (p. 194). These definitions make it clear that self-efficacy is specific, rather than global, and must be understood and assessed as such. Also, self-efficacy must be measured with strict regard to the related task. And by defining self-efficacy in this way, Pajares and Miller more clearly differentiated self-efficacy from self-esteem. Although self-efficacy can be generalized (i.e., generalized self-efficacy) so that it is defined as the ability to succeed in new tasks or challenges, it is still distinguished from self-esteem in that self-esteem is tied to self-worth, whereas self-efficacy is linked to perceived abilities.

Before moving forward with the discussion of academic self-efficacy, it is essential to discuss academic self-efficacy in relation to another construct that is also at times used interchangeably with self-efficacy: self-concept. Some authors have distinguished academic self-concept and academic self-efficacy. Self-concept is described as the belief that one is broadly competent at academics, whereas academic self-efficacy is defined as the belief that one is skilled in performing specific academic tasks (Bong & Clark, 1999; Bong & Skaalvik, 2003; Scherrer & Preckel, 2019). However, many others have noted a strong correlation between the constructs (Ferla, Valcke, & Cai, 2009; Trautwein & Möller, 2016), and that due to their strong correlation, research regarding academic self-concept can be used when examining academic

self-efficacy. Indeed, academic self-concept and academic self-efficacy are often used interchangeably in the literature.

Self-Efficacy and Achievement

Relation between self-efficacy and achievement.

Self-efficacy is often discussed with academic achievement. Several meta-analyses have been conducted to examine the relationship between academic achievement and academic self-efficacy (Richardson, Abraham & Bond, 2012; Scherrer & Preckel, 2019). Multon, Brown & Lent (1991) conducted one such meta-analysis to examine the facilitating relationship of self-efficacy beliefs to academic performance. They reviewed 38 studies, 25 of the 38 studies were experimental studies of self-efficacy treatments, and 13 were non-experimental in design. Effect sizes were calculated and used to compute an aggregate effect size. The aggregate effect size estimate was $r = .38$. Across various types of student samples, designs, and criterion measures, self-efficacy beliefs accounted for approximately 14% of the variance in students' academic performance. There was, however, significant heterogeneity in the studies, so the correlation depends on specific study characteristics.

Bong and Clark (1999) defined academic self-efficacy as "student's beliefs about their capability to succeed in specific academic pursuits" (p. 141). In this context, self-efficacy may differ across different academic content areas. Indeed, Bong (1998) studied the generality of academic self-efficacy judgments among 588 high school students. In the study, participants rated their self-efficacy for solving 42 different types of problems in various academic subjects, one of which was algebra. Bong (1998) found that students rated their strengths comparably as they perceived the similarity between the kinds of problems being presented. For example, students who believed themselves to be efficacious in algebra were more likely to believe

themselves to be efficacious in other mathematics tasks, but not as likely to consider themselves to be efficacious in other areas, such as literature. The students in Bong and Clark's (1999) study generalized their beliefs based on associations that they developed from problem-solving experiences. These beliefs could be generated based solely on the perception of task abilities, even if the student has not experienced that task.

Self-efficacy and achievement theory.

Marsh's (1990) internal/external model (I/E model) is a relevant theoretical framework for further understanding of how students develop academic self-efficacy. According to the I/E model, academic self-efficacy in a particular school subject is formed in relation to an *external* (social comparison) reference: Students compare their self-perceived performances in a specific school subject with the perceived achievements of other students in the same school subject. It is also formed in relation to an *internal* reference: Students compare their performances in a particular school subject with their performances in other school subjects. Hence, students may have a favorable mathematical self-efficacy, if mathematics is their best subject, even if they are not particularly good at mathematics relative to other students. This model explains why self-efficacy is more likely to vary across achievement areas, despite the strong correlation of achievement across academic fields.

The I/E model was examined in a meta-analysis by Möller, Pohlmann, Köller & Marsh (2009). The meta-analysis included 69 studies ($N=125,308$) from the United States, Asia, Australia, and Europe. The results from the meta-analysis supported the I/E model and demonstrated that the model held across age, gender, and country of origin. Correlations between self-efficacy and verbal or mathematics achievement within the studies used in the meta-analysis ranged from $r = .17$ (Marsh, Smith, & Barnes, 1984) to $r = .73$ (Skaalvik & Rankin, 1995), so

obviously the correlation depends to some extent on unique characteristics of each study. Mathematical and verbal achievement scores were moderately highly correlated overall ($r = .67$), but the correlation between mathematical and verbal self-efficacy was weak ($r = .10$). Positive paths from achievement to the corresponding self-efficacy ($\beta = .61$ for math, $\beta = .49$ for verbal) and negative paths from achievement in one subject to self-efficacy in the other subject ($\beta = -.21$ from mathematics achievement on verbal self-concept, $\beta = -.27$ from verbal achievement to mathematical self-concept) were found. They found a moderate aggregate correlation between mathematical self-efficacy and mathematics achievement $r = .37$ when mathematics achievement was measured using test scores, and $r = .50$ when mathematics achievement was measured using grades, so the type of achievement measure may affect the relationship. Overall, however, the findings support the I/E model, as the achievement scores in mathematical and verbal domains were moderately highly correlated, as were those academic areas to their respective areas of self-efficacy. However, verbal self-efficacy was not correlated with math self-efficacy.

The Eccles, Adler, Futterman, Kaczala, and Meece (1983) expectancy-value model of achievement-related choices is another useful theoretical framework for understanding a theoretical reciprocal relationship between academic self-efficacy and academic achievement. According to this model, students' decisions to persist in taking coursework is determined by their assessments of the likelihood of success in, and the relative value that they assign to, the options perceived to be available. (Eccles, 2009). More self-efficacious students participate more readily, work harder, persist longer, and have fewer adverse emotional reactions when they encounter difficulties on tasks than do those who doubt their capabilities. Regarding the choice of activities, self-efficacious students undertake difficult and challenging tasks more readily than

do inefficacious students. Self-efficacy beliefs are predictive of two measures of students' effort: rate of performance and expenditure of energy. This prediction is mediated by the choice to take challenging coursework and complete practice in challenging tasks. In turn, it has been hypothesized that self-efficacy influences performance, while performance in an area also reinforces self-efficacy.

Growth in academic self-efficacy in school-age students.

When academic, or specific aspects of academic self-efficacy, have been examined longitudinally, there has generally been an observed a small mean-level decrease across a student's development and through their academic career (Anderman, Maehr, & Midgley, 1999; Bong & Skaalvik, 2003; Pajares & Graham, 1999; Spinath & Spinath, 2005; Wigfield et al., 1997). In one such study, Pajares & Graham (1999) measured student mathematical self-efficacy on a 1-8 scale in the fall, and then again in the spring, during the first year of middle school. In female students, they observed a mean decrease from 7.0 to 6.3, and in male students, there was a mean decrease from 7.2 to 6.4 across the school year. However, other studies have shown that levels of self-efficacy either remain the same or show slight increases over time, especially over multiple years. One such study by Shell, Colvin, & Bruning (1995) measured reading and writing self-efficacy from 4th to 10th grade, at three time points, on a 1 to 5 scale. Students showed average reading self-efficacy ratings of 4.01, 3.98 and 3.99, in 4th, 7th, and 10th grade, respectively. Thus, mean reading self-efficacy ratings were the same over time. However, in the same study, students showed average writing self-efficacy ratings of 4.31, 4.53, and 4.70, in 4th, 7th, and 10th grade, respectively, which showed mean writing self-efficacy ratings increase slightly over time. The mixed results regarding the mean-level development might be explained by the different grade levels of the samples, different scales of self-efficacy used, and different

types of self-efficacy assessed (i.e., math, reading, and writing). Alternatively, it may be that on average self-efficacy decreases slightly over a school year, but may not change or slightly increase across school years. Nevertheless, because changes were so small, it may be interpreted that, on average, efficacy does not increase or decrease in a substantially meaningful way over time (Scherrer & Preckel, 2019).

Co-development of self-efficacy and achievement.

Of particular interest to the present study is the co-development of self-efficacy and achievement, thus the relation between the two over time. Theoretically, there are four possible ways in which self-efficacy and achievement could be related to one another: (1) self-efficacy and achievement are not related, (2) self-efficacy primarily affects later achievement, or the self-enhancement model, (3) achievement primarily affects self-efficacy, or the skill development model, (4) self-efficacy and achievement affect each other, or a reciprocal model (Ganley & Lubienski, 2016; Ganley & Vasilyeva, 2011; Huang, 2011; Valentine, DuBois, & Cooper, 2004). The best support is found for the skill development model, the self-enhancement model, and the reciprocal model, with limited support for the model indicating no relationship. Hence, there is a longitudinal relationship, but it is not clear which model best describes it (Huang, 2011).

Valentine, DuBois, & Cooper (2004) conducted a meta-analytic review of 56 longitudinal studies examining the relation between self-beliefs, including self-efficacy, and academic achievement. They found that, after controlling for prior academic achievement, the average effect of prior self-efficacy on subsequent academic achievement was small ($\beta = .10$). The average effect of previous academic achievement on subsequent self-efficacy was also small ($\beta = .08$), for both grades and test scores. The finding supports the reciprocal model: Academic achievement has a positive impact on subsequent academic self-efficacy, and academic self-

efficacy has a positive effect on subsequent academic achievement. However, the effect size for both longitudinal relations was small.

In a similar meta-analysis, Huang (2011) examined the relationship between self-efficacy (the variable was labeled self-concept in the study but had a similar definition to self-efficacy as understood here) and academic achievement, using 39 independent and longitudinal samples. Path analysis procedures were used to determine the best model for understanding the relationship. The mean observed correlations ranged from .20 to .27 between prior self-efficacy and subsequent academic achievement and from .19 to .25 between prior academic achievement and subsequent self-efficacy. The findings also support the reciprocal model, showing that academic achievement has a positive impact on academic self-efficacy and academic self-efficacy has a positive impact on academic achievement.

The effect sizes reported in Huang were larger than those reported by Valentine et al., (2004), but that is because Valentine and colleagues summarized regression coefficients from models that included other potential predictors in the models. In both the Valentine et al. (2004) and the Huang (2011) study, the specificity of the measure of self-efficacy was found to be a significant moderating factor in the relation between both prior self-efficacy and subsequent academic achievement and previous academic achievement and subsequent self-efficacy: More specific measures of self-efficacy have stronger relationships with particular areas of achievement. Of particular interest to this study is the longitudinal relation between mathematical self-efficacy and mathematics achievement. So, it can be seen that the longitudinal relationships between mathematical self-efficacy and mathematics achievement in this study may be stronger than those reported in the Valentine et al. (2004) study.

Mathematical Self-Efficacy and Mathematics Achievement

Self-efficacy is understood as an internal or cognitive assessment or judgment of specific capabilities or competencies to address a particular task or similar tasks at hand (Pajares & Miller, 1994). Academic self-efficacy is an internal assessment of competence in academics. Self-efficacy can be defined even more precisely. For example, mathematical self-efficacy is an internal assessment of ability in mathematics. It differs from academic self-efficacy in that it is specific to mathematics, and for example, not an internal evaluation of proficiency in reading. The literature conclusively indicates that there is a positive relationship between mathematical self-efficacy and mathematics achievement (Möller, et al., 2009). It has been shown to have a stronger relationship with STEM achievement compared with other variables commonly associated with mathematical performance, such as perceived usefulness of mathematics, prior experience with mathematics, or gender (Pajares & Miller, 1994).

The High School Longitudinal Study of 2009 is the dataset used in the present analysis. The researchers established a self-efficacy scale, which was constructed using responses to the following Likert scale items: "Confident s/he can do an excellent job on tests in current mathematical class," "Understands most difficult textbook material in current mathematical class," "Certain s/he can master skills being taught in current mathematical class," "Confident s/he can do an excellent job on current mathematical class assignments" (Ingels et al., 2013). These questions capture several components of mathematical self-efficacy identified in research.

Mathematical self-efficacy beliefs have been shown to influence critical components of academic motivation, such as choice of activities, level of effort, persistence, and emotional reactions, and thus influence mathematics achievement. For example, Schunk and colleagues (1987) found a moderate positive correlation ($r = .51$) between perceived mathematical self-

efficacy and students' rate of solution of arithmetic problems. This study also showed that students' mathematical self-efficacy beliefs were predictive of their choice of engaging in subtraction problems rather than in a non-mathematical task: The higher the student's sense of mathematical self-efficacy, the more likely they chose an arithmetic activity.

Mathematical self-efficacy and mathematics achievement longitudinal research.

Although there has been considerable study of the broader constructs of academic self-efficacy and achievement, there has been substantially less study of mathematical self-efficacy and mathematics achievement, especially in U.S. student populations (Valentine, DuBois, & Cooper, 2004). Further, much of the research of this specific relation has not been conducted on a large sample using longitudinal data (i.e., repeated measurements of both mathematical self-efficacy and mathematics achievement), and therefore, does not consider prior achievement and self-efficacy, limiting conclusions about how the correlation between mathematical self-efficacy and mathematics achievement arises (Huang, 2011). Analyzing a large sample with longitudinal data on achievement and self-efficacy collected at multiple time points can shed light on the relation between the two, and possibly help to tease apart the direction of the connections.

One longitudinal study was conducted by Skaalvik and Valås (1999). The study consisted of three cohorts made up of Norwegian students in 3rd, 6th and 8th grade, totaling 1,005 students. Students were administered measures twice, within each grade level, at the beginning and end of each grade. Variables included were mathematics achievement (measured by teacher ratings on a 6 point scale), mathematical motivation (measured by student ratings in mathematical interest and mathematical investment on a 5 point scale), and mathematical self-efficacy (the self-efficacy variable was called self-concept, though was measured using similar items to the self-efficacy variable here). The mathematics achievement and mathematical motivation measures for

all three cohorts showed moderate to very high stability from time one to time two. The stability of mathematical self-efficacy, however, increased dramatically with the age of the students. Very low stability across the school year was observed in 3rd grade ($r = .20$); it increased with age ($r = .48$ and $r = .70$ in 6th and 8th grade, respectively). In all cohorts, the results were consistent with the skill-development model, that is, the view that achievement affects subsequent self-efficacy ($\beta = .26, .35, .15$ in 3rd, 6th, and 8th grade, respectively). No evidence was found that self-efficacy affects subsequent achievement (self-enhancement model) or for the reciprocal model.

Interestingly, in the two oldest cohorts, motivation was influenced by the previous achievement. However, there was no evidence that self-efficacy affects either subsequent motivation or achievement. One limitation of the study was that teacher rating of achievement was used and not an objective achievement measure.

Ganley and Lubienski (2016) conducted another notable, recent study that examined repeated measures of both mathematical self-efficacy and mathematics achievement. This study examined the gender differences in the relationship between mathematical self-efficacy, mathematical interest, and mathematics achievement. In this study, the researchers explored gender-related patterns in mathematical related self-views and achievement, following students from 3rd to 8th grade, with data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999. Results indicated that gender differences in mathematical self-efficacy (defined as confidence, but similar to how self-efficacy is used here) were more extensive than gender differences in interest and achievement in middle school. Longitudinal cross-lag panel models showed that mathematics achievement was a predictor of subsequent mathematical self-efficacy ($\beta = .24-.27$) through elementary and middle school, even after controlling for prior levels of mathematical self-efficacy. Again, these findings support the skill development model. There

was not a reciprocal relationship because self-efficacy did not predict later achievement in a meaningful way, after controlling for prior levels. These relations were the same for male and female students. Further, mathematical self-efficacy did show a small mean increase across the grade levels measured, from 3rd to 8th grade.

The studies both supported the skill development model, where mathematics achievement affects later mathematical self-efficacy. The reciprocal model with mathematical self-efficacy and mathematics achievement was not supported in either the Skaalvik & Valås (1999) or the Ganley & Lubienski (2016) study, but these studies only followed students from elementary to middle school. Further, in both studies, they noted an increase in the role of some of the mathematical self-worth (motivation and confidence), as the student's aged. If these patterns continue to demonstrate an increase with age, the reciprocal model that includes mathematical self-efficacy may also be supported for students at the high school age, which will be examined in the present proposed study.

Self-enhancement model

In contrast to the above studies, the self-enhancement model, showing that mathematical self-efficacy affects subsequent mathematics achievement, was supported in a study by Marsh and colleagues (2005). This study examined two nationally representative samples of German 7th grade students (Study 1: $n = 5,649$, Study 2: $n = 2,264$). Prior self-efficacy (labeled self-concept but defined similarly to self-efficacy in this study) significantly affected subsequent school grades (Study 1 $\beta = .28$, Study 2 $\beta = .27$) and standardized test scores (Study 1 $\beta = .15$, Study 2 $\beta = .21$). Prior grades and test scores showed a minimal ($r = .03$ and $.08$, $r = .06$ and $.03$, respectively) correlation with self-efficacy, but each was only statistically significant in one of the two studies. There were also significant standardized mean differences in gender, using a

latent mean difference analysis, controlling for time one responses, male students showed higher mathematical self-concept ($d=.12$), mathematical interest ($d=.10$), but showed lower mathematical grades (.08), and test scores (.06). For mathematics test scores, male students scored higher than female students did at time one, but their advantage was much smaller at time two so that after controlling for time one scores female students actually did slightly better than male students at time two. For mathematics grades, there were no significant differences between male and female students at time one, and female students performed marginally better than did male students at time two. Despite the mean differences, the relationship of the variables was not affected by gender as a moderator.

Reciprocal model

Finally, a study by Pinxten et al. (2014) supported the reciprocal model. Structural equation modeling was used, in a five-wave design with a cohort of 4,724 Flemish students in grades 3-7, to examine the longitudinal reciprocal interrelations between mathematical competency beliefs, mathematical enjoyment, mathematical effort, and mathematics achievement. The researchers observed that mathematical self-efficacy (there referred to as mathematical competency beliefs) predicted mathematics achievement in all measured grades ($\beta=.10, .13, .14$ in 4th, 5th and 6th grade respectively). Moreover, mathematics achievement predicted mathematical self-efficacy in all measured grades ($\beta=.50, .21, .22, \text{ and } .16$ in 3rd, 4th, 5th, and 6th grade respectively). These findings indicate a reciprocal, mutually beneficial, relationship between mathematical self-efficacy and mathematics achievement. Additionally, the researchers found a positive relationship between past achievement and later interest but a negative relationship between previous mathematical interest and later achievement when mathematical self-efficacy was included in the model. These results suggest that the reciprocal

effects model may be a better representation of the relationship between self-efficacy and achievement than interest and achievement when they are considered in a model together.

Thus, although there has been ample longitudinal study of the broader construct of academic self-efficacy and achievement, there is notably less study of mathematical self-efficacy and mathematics achievement using longitudinal data, especially in U.S. student populations (Valentine, DuBois, & Cooper, 2004). Further, much of the research of this specific relation has not been conducted on a large sample using longitudinal data (i.e., repeated measurements of both mathematical self-efficacy and mathematics achievement), and therefore, does not consider prior achievement and self-efficacy, limiting conclusions about how the correlation between mathematical self-efficacy and mathematics achievement arises (Huang, 2011). Finally, much of the longitudinal research has examined self-concept or competence. While these terms are highly related to mathematical self-efficacy, the studies may lack the precision of measurement of a study focused on the construct.

There have been several pertinent studies that can provide context for the present inquiry, all indicating that there is a longitudinal relationship between mathematics achievement and mathematical self-efficacy. However, the relevant studies support different conclusions about what the nature of that longitudinal relationship. With evidence for the self-enhancement model (Marsh et al., 2005), the skill development model (Ganley & Lubienski, 2016; Skaalvik & Valås, 1999), and the reciprocal model (Huang, 2011; Pinxten et al., 2014; Valentine et al., 2004), the relationship between mathematical self-efficacy and mathematics achievement is uncertain. Also, the studies reported here are primarily on international populations (Marsh et al., 2005; Pinxten et al., 2014; Skaalvik & Valås, 1999), students in elementary and middle school (Ganley & Lubienski, 2016; Marsh et al., 2005; Pinxten et al., 2014; Pomerantz, Altermatt, & Saxon,

2002), and in college-age populations (Marshman et al., 2018; Robnett, Chemers, & Zurbriggen, 2015). More information is needed about the nature of these relationships in high school students in the United States.

Mathematical Identity

One of the critical questions of identity theory is how one's identity develops. Erikson (1968) put forth one foundational theory explaining how identity evolves as people age. Erikson contended that people face a variety of developmental tasks throughout their lifespan. In adolescence, the construction of a synthesized sense of identity becomes prominent with essential implications for personal and social adjustment. These adolescent identities are formulated in relation to academic, educational, and relational domains (Klimstra et al., 2010). As with self-efficacy, identity can be viewed through an adjustable lens. Identity can be interpreted via the big picture, how a person broadly understands and defines themselves, or it can be narrowed to a specific area (Stinson & Bullock, 2012). The focus of the present study is on mathematical identity.

Broadly, mathematical identity refers to a person's beliefs, attitudes, emotions, and dispositions about mathematics. It involves the ways students think about themselves concerning mathematics. Some theorize that identity develops through social processes and shared experiences that promote a sense of self and meaning. Martin (2006) stated, "A mathematics identity encompasses a person's self-understandings as well as how they are constructed by others in the context of doing mathematics" (p. 206). Students learn to engage in mathematics practices through school and other academic communities and develop a sense of self in relation to those practices and academic communities in which they learn. Two pillars support a positive mathematical identity: the belief that you can do mathematics and the belief that you belong in

mathematics contexts (such as a mathematics classroom). Having a sense of identity and community fosters students' engagement in mathematics, persistence in the face of challenges, and adoption of effective academic behaviors (Froschl & Sprung, 2016).

The development of mathematics-based identities has been widely studied in recent years, yet the method of measurement, the discipline of study, and construct definitions have varied widely. Two critical systematic literature reviews have attempted to clarify this definition problem. Darragh (2016) conducted a systematic literature review of 188 articles from 85 different journals on the theoretical underpinnings of the concept of mathematical identity and its use in mathematics education research. She concluded that identity might be seen as an action and fit within a sociological frame, or it may be seen as an acquisition, matching within a psychological framing (Darragh, 2016). In other words, identity can be seen sociologically within the context of society and how one fits within a group, or it can be seen psychologically as something one develops internally, with a variety of influences from learning, models, and experiences.

The other recent systematic literature review that has attempted to take on this issue was completed by Radovic, Black, Williams, and Salas (2018). These authors reviewed 69 research papers to summarize how the concept of identity has been employed in mathematics research. They found that identity has been conceptually defined under three main dimensions: social/subjective, enacted/representational, and change/stability. Additionally, they determined that identity has been implemented within the three dimensions in five ways: a). Identity as an individual attribute, b). Identity as narrative c). Identity as a relationship with a specific practice d). Identity as a way of acting and e). Identity as afforded and constrained by local traditions. Thus, while there is a broad understanding of identity, the definition is not clear, and it is

subjective. Despite agreement about the relevance of the concept of identity, subjectivity in the interpretation makes it challenging to study as a single construct.

The challenge in developing a foundational literature review on a singular definition of identity is particularly true because there are a wide variety of disciplines, using a wide array of methods that engage in identity research, outside of school psychology (Radovic et al., 2018). There are apparent epistemological differences between these perspectives. Yet, Grootenboer, Zevenbergen, and Chinnappan (2008) suggested that the plurality of theoretical lenses provide a more productive and more comprehensive understanding of the issues of mathematical identity development. In the present dissertation, mathematical identity is operationally defined as the self-definition of a person thinking of themselves, and perceiving that others think of them as a "math person." The researchers that constructed the scale from the High School Longitudinal Study of 2009 used responses to the following Likert scale items: "Student sees her/himself as a math person," "Others see the student as a math person," (Ingels et al., 2013). This definition encompasses some components of the broader definitions of mathematical identity that other researchers have forwarded.

Mathematical Identity and Mathematics Achievement

Mathematical identity may be a useful tool in understanding mathematical engagement. The processes related to students' mathematical identity development are central to students' mathematics content acquisition, such as participation in mathematical classes and completion of mathematics homework. The importance of understanding student beliefs in more specific contexts is further supported by the structured relationships discovered between identity, goals, and practice (Esmonde, 2009). Students seem to develop their sense of identity through participation and achievement in related tasks. As they set and achieve goals within a job, they

develop skills and mastery, and then that skill becomes an identity. In other words, students create a sense of their relative competencies (self-efficacy), and those competencies influence the values that identity is based on during the adolescent period (Wigfield & Wagner, 2005). Indeed, recently some researchers have begun to forward the idea that the goal of a mathematics teacher is to build their students' mathematical identity rather than merely instruct them in mathematics achievement (Grootenboer, Zevenbergen & Chinnappan, 2008). Nevertheless, there is limited empirical evidence that there is a link between mathematical identity and mathematics achievement.

Mathematical identity as it relates to mathematics achievement has been primarily studied using qualitative measures, this is, in large part because of the way that most researchers define identity as a changing construct, developed through the ways that people communicate with themselves or others, as described through narrative (McGee & Martin, 2011; Radovic et al., 2018). As Darragh (2016), in her systemic literature review, explained, often, "identity is seen as a complex concept requiring detailed descriptions of individuals rather than generic findings of larger groups." Most (58%) of the articles included in her literature review reported on fewer than 10 participants and employed qualitative methods such as interviews, observations, video analysis, autobiographies, and document analysis. Even when quantitative methods have been used for studies of mathematical identity, they tend to be from small sample sets, focused on a particular group or population, or done within the context of a more extensive STEM study, often related to physics or engineering (Kost, Pollock, & Finkelstein, 2009). Neither the Radovic, et al. (2018) nor the Darragh (2016) systematic literature reviews contained a study similar to the one being presented here, using a modern, large-sample data set.

Williams, Burt, and Hilton (2016) conducted a study using data from the National Education Longitudinal Study of 1988 for a historical examination of various student affect variables that relate to mathematics achievement in high school. A very small, but statistically significant correlation between mathematical identity and mathematics achievement was found ($r = .07$). This finding suggests that there may be a minimal relationship between a student's mathematics achievement and mathematical identity.

While there has been limited quantitative research of mathematical identity and its relationship to mathematics achievement, there has been some related research done examining the relationship between physics identity and performance in physics classes. Kost, Pollock, and Finkelstein (2009) conducted one such study. These researchers surveyed approximately 900 students across two semesters of an introductory physics course at the University of Colorado at Boulder. Constructs evaluated in the survey were physics self-efficacy and physics identity. A moderate to moderately strong correlation was found between all variables. For physics identity, a correlation of .32 was found for course grade and .36 for final exam score. For physics self-efficacy, a correlation of .50 was found for course grade and .58 for final exam score. The correlation between physics self-efficacy and physics identity was not reported.

There were also some interesting findings related to gender in the Kost, Pollock, and Finkelstein (2009) study. There was not a significant gender difference in student grades for the course. However, female students rated themselves significantly lower on the items measuring physics self-efficacy, even when controlling for exam scores. Female students also rated themselves significantly lower on some, though not all, of the items measuring physics identity. The researchers also administered the Force and Motion Conceptual Evaluation (FMCE), which is a tool used to evaluate conceptual knowledge in physics; the instrument was administered both

before and after the course. Female students scored significantly lower on both the pretest and posttest, suggesting a gender gap in achievement that may not have been shown in course grades or test scores. Despite this, the correlation between FMCE performance and physics self-efficacy and physics identity was moderate and only slightly weaker than for course grades. For identity, a correlation of .24 was found for both the pre and posttest. For self-efficacy, a correlation of .44 was found for the pretest and .38 for the posttest.

The remaining quantitative studies that examined mathematical identity and mathematics achievement used the High School Longitudinal Study of 2009 (HSL:09), which is being used here (Middleton 2013; Middleton, Mangu, & Lee, 2019). Despite the use of the dataset in those studies, each one is distinct from the present study. The most related past study using HSL:09 was conducted by Middleton (2013), which performed a path analysis using the Base Year data from the HSL:09 dataset. It was not clear if mathematical identity was associated with achievement in the study. Still, it was related to mathematical self-efficacy, but again the relations were not evident in the model, and correlations were not reported. Also, this study did not use data from the follow-up, so no longitudinal analysis was conducted. Another highly related study was done by Middleton, Mangu, and Lee (2019), which examined time one and time two data to determine how mathematical self-efficacy, mathematical identity, mathematical interest, and mathematical utility predict STEM career intentions and motivation. The variables relationship with mathematics achievement was not studied.

In summary, there were few quantitative studies regarding the relation between mathematical identity and mathematics achievement. There is some evidence that suggests a small relationship, yet the theory supports the possibility of a stronger relationship than is illustrated by the studies presented here. It may be that mathematical identity acts as a suppressor

in that it enhances the relation between mathematical self-efficacy and mathematics achievement. Mathematical identity and mathematical self-efficacy may be strongly correlated (Huang, Cribbs & Piatek-Jimenez, 2016), mathematical self-efficacy and mathematics achievement are moderately correlated (Schunk et al., 1987), but mathematical identity and mathematics achievement are nearly independent (Williams, Burt & Hilton, 2016).

Mathematical Self-Efficacy and Mathematical Identity

Science, technology, engineering, and mathematical educators and researchers have recognized the importance of examining affective components of learning. Trujillo & Tanner (2014) noted that “monitoring students' self-efficacy, sense of belonging, and science identity” are important goals for future researchers as these are “areas that may well improve educator effectiveness.” This literature review has examined self-efficacy—the set of beliefs that one is capable of performing a task; and mathematical identity—the extent to which a person is recognized or recognizes himself or herself as a “math person.”

As there have been exceedingly few large sample studies examining mathematical identity, the number of studies examining mathematical identity and mathematical self-efficacy is sparse. Huang, Cribbs, and Piatek-Jimenez (2016) surveyed 323 undergraduate students at a large public university in the Northeastern region of the United States. The researchers examined mathematical self-efficacy and mathematical identity as they relate to STEM career interest. A large correlation between mathematical self-efficacy and mathematical identity was found ($r = .81$). This finding suggests that mathematical self-efficacy and mathematical identity are strongly related but distinct concepts. In another study with high school students, the relation between the two was shown in a path model. After controlling utility and interest in mathematics,

mathematical self-efficacy was strongly related ($\beta = .58$) to mathematical identity (Middleton 2013).

In another related study of science identity and science self-efficacy, Robnett, Chemers, and Zurbriggen (2015) conducted a 2-year longitudinal study of 251 undergraduates recruited from colleges and universities across the United States. Data was gathered on the students' research participation, science self-efficacy, and science identity at three-time points during the two years. Interestingly, identity as a scientist at time one predicted science self-efficacy at time two ($r = .14$), but science identity at time two did not predict science self-efficacy at time three. Likewise, science self-efficacy at time two predicted science identity at time three ($r = .17$), but science self-efficacy at time one did not predict science identity at time two. These results support a possible reciprocal model, though it appears that the reciprocal relationship occurs over time, not co-occurring. Follow-up tests examined whether participant gender moderated paths in the model, but it did not. Further, the model explained similar amounts of variance in women's and men's time three identity as a scientist (R^2 women = .28; R^2 men = .22).

Also, it can be intuitively observed, from the theories of self-efficacy and identity that there is a likely relationship between mathematics achievement, mathematical identity, and mathematical self-efficacy. Mathematical self-efficacy and mathematical identity are essential factors in motivation towards completing mathematics tasks (Eccles, 2009; Esmonde, 2009). Further, the experiences a person has are critical in the development of these self-beliefs. Concerning mathematical self-efficacy and mathematical identity, prior experiences in domain-specific academic endeavors shape student's perceptions of competence in that area, or their self-efficacy, and their understanding of themselves or their identity beliefs. These beliefs include

assumptions about performance, knowledge, constraints, and opportunities, which are components of prior mathematics experience (Lent, Brown, & Gore, 1997).

Gender's Relationship with Mathematical Self-Efficacy, Mathematical Identity & Mathematics Achievement

Research has shown that female students have both lower mathematical self-efficacy (Brainard & Carlin, 1998; Else-Quest et al., 2010) and lower identification with STEM fields than male students (Hazari et al., 2010). This finding persists despite consistent evidence that, when controlling for access to education, female students tend to achieve at approximately the same level as male students on most mathematics tests and grades (Else-Quest et al., 2010). Lower levels of mathematical self-efficacy for female students have been documented even in samples in which female students obtain higher mathematics grades than male students (Pomerantz et al., 2002). Lower levels of mathematical self-efficacy and mathematical identity are associated with a lower probability of engaging in STEM experiences, choosing STEM careers, and pursuing advanced STEM education (Middleton, Mangu, & Lee, 2019). This has produced a gender gap in representation in advanced STEM education and STEM career fields (Botella, Rueda, López-Iñesta, & Marzal, 2019).

Brainard and Carlin (1998) conducted a longitudinal study of the persistence of women in undergraduate science and engineering majors at the University of Washington between 1991 and 1996. Using annual surveys of female science and engineering majors, they tracked several cohorts of students through the university to try to understand why some women leave science or engineering majors, and others stay. Students in this study reported isolation, a lack of belongingness, intimidation, and a lack of self-efficacy as they progressed through their major program. Interestingly, there was no difference in performance (as measured by GPA) between

those women who remained and those who transferred out of STEM programs. It seems then that those women who report lower self-efficacy and identity in STEM areas may be more likely to change majors out of science and engineering ratings than their peers who stayed even though both groups had the same GPA.

Else-Quest et al. (2010) meta-analyzed two major international data sets—the 2003 Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA), representing 493,495 students, 14–16 years of age—to estimate the magnitude of gender differences in mathematics achievement, attitudes and affect across 69 nations throughout the developed world. The weighted mean effect sizes in mathematics achievement were very small (TIMSS: $d = -.01$, PISA: $d = .11$). When controlling for educational access, achievement was consistent across genders. However, male students showed moderately higher self-efficacy in both datasets. On the TIMSS, the weighted mean effect size for self-confidence in mathematics was a .15, while on the PISA, the weighted mean effect size for self-concept and self-efficacy were both .33.

In another study, Pomerantz and colleagues (2002) evaluated 932 elementary age school children on their self-perception of confidence (self-efficacy) in math, language arts, social studies, and science. The researchers also obtained the students' grades in those academic areas. The mean effect size in all academic areas showed that female students achieved higher than male students (science: $d = .15$, math: $d = .14$). However, female students rated themselves lower on their self-perception of their abilities. The mean effect size of self-efficacy in mathematics was $-.26$, and the mean effect size of self-efficacy in science was $-.22$. Again, there is a gender gap in self-efficacy in both the elementary and adolescent students.

As none of the above studies are longitudinal, they do not allow researchers to control for previous grades when examining the effect size of the gender gap in self-efficacy. Marshman and colleagues (2018) conducted a study of about 2000 college students in physics I and physics II courses, at the University of Pittsburgh. Students were measured in their self-efficacy and achievement twice to answer the question, “What is the self-efficacy of female and male students throughout a two-semester physics course sequence when students’ course performance is accounted for?” Female student’s grades were relatively consistent with male students, and there was no interaction effect between gender and grade. However, female students had significantly lower self-efficacy compared to their male counterparts, Cohen’s d was .60 or higher for physics I and physics II, in all grade groups (A, B, C). In fact, female students receiving A’s had similar self-efficacy as male students receiving C’s. In other words, the effect of gender on self-efficacy did not change based on the course grade. The researchers also performed an ANCOVA to determine what portion of the self-efficacy gap can be attributed to performance. They found that only a small part of the gender gap in self-efficacy might be attributable to differences in performance. The gap seems to mainly come from biased perceptions: 87% in physics I and 95% in physics II.

There have been fewer studies of a potential gender gap in mathematical identity done on large groups of participants. Hazari and colleagues (2010), however, examined the responses of 3,829 students from 34 randomly selected US colleges/universities to evaluate the relationship between physics identity and career choice and found that male students had a significantly higher physics identity indicator than their female counterparts by about half a standard deviation. In a follow-up study, using the Kruskal-Wallis tests for gender and race/ethnicity comparisons, Hazari and colleagues (2013) found that male students, on average, perceive

themselves as a physics person more than female students do, regardless of race/ethnicity. They used data from the Persistence Research in Science and Engineering (PRiSE) project, which surveyed 7,505 students from 40 colleges and universities across the United States about their backgrounds, high school science experiences, and science attitudes.

Yee & Eccles (1988) conducted a study of parent perceptions and attributions of children's math achievement. In the study, they examined the parents of 48 junior high male students and female students of high, average, and low math ability groups. They found that girls reported lower estimations of their math ability and expressed more negative attitudes about math than boys, despite equivalent performance in grades. Their parents showed the same sex-typed bias. Interestingly, parent's math-related perceptions and attributions varied with their child's gender. Parents credited daughters with more effort than sons, and sons with more talent than daughters for successful math performances. From this study, we can begin to theorize that female students may be less influenced by their mathematics achievements in their development of mathematical self-efficacy and mathematical identity. Female students receive feedback telling them that their accomplishments are not associated with their own characteristics (such as mathematical self-efficacy and mathematical identity) but rather due to their hard work. In contrast, boys may receive feedback that causes them to internalize successes, tying those successes to their inherent abilities, which then become connected to their level of mathematical self-efficacy and mathematical identity.

Also, consider Marsh's (1990) I/E model. According to the I/E model, academic self-efficacy in a particular school subject is formed in relation to an *external* (social comparison) reference: Students compare their self-perceived performances in a specific school subject with the perceived achievements of other students in the same school subject. It is also formed in

relation to an *internal* reference: Students compare their performances in a particular school subject with their performances in other school subjects. While male and female student's internal references may act similarly, female students may rely more on social comparisons to develop their sense of mathematical identity and mathematical self-efficacy than male students, which has been observed in studies examining the role of stereotype threat in academic performance (Marx, Stapel, & Muller, 2005). As female students may rely more on social comparisons, but receive less positive external feedback due to the narrative that their success is due to hard work, not internal characteristics (Yee & Eccles, 1988), this may push them to be even less likely to show strong mathematical self-efficacy and identity in comparison to male students, despite equivalent grades.

Finally, it is essential to note here that a binary understanding of sex and gender merely to male and female limits the conclusions available from this study. Given how the sex/gender variable was assessed in the dataset, students were forced to identify as male or female while completing the survey. It is now recognized that it is crucial to allow students to identify their sex and gender along a spectrum, to include students who are intersex, transgender, non-binary, or otherwise do not identify as male or female. Despite this limitation within the dataset, there is still value in studying students who identify as male or female, as this does provide some insights, even if limited, into differences that may be caused by the different social experiences of these groups of students (Hyde, Bigler, Joel, Tate, & Van Anders, 2019).

In summary, women generally demonstrate similar or even stronger STEM grades and test scores to men (Else-Quest et al., 2010; Ganley & Lubienski, 2016; Marsh et al., 2005; Pomerantz et al., 2002). While these findings support the second hypothesis, that although male students are predicted to show lower mathematical identity and self-efficacy than female

students, the temporal relation among mathematical self-efficacy, mathematical identity, and mathematics achievement will not differ between male and female students, there is still value in examining the research question. The lack of difference observed stands in contrast with popular theory, which suggests an achievement gap, and instead indicates that distinctions in belongingness, identity, and self-efficacy may be a more influential factor in a female student's STEM experience (Froschl & Sprung, 2016). Lower levels of mathematical self-efficacy and mathematical identity are associated with a lower probability of engaging in STEM experiences, choosing STEM careers, and pursuing advanced STEM education (Middleton, Mangu, & Lee, 2019). This has produced a gender gap in representation in advanced STEM education and STEM career fields (Botella, Rueda, López-Iñesta, & Marzal, 2019). Further, it is anticipated that if gender does moderate the relationship, it is predicted that self-beliefs will be affected more strongly by achievement in male students than in female students. Therefore, in an attempt to better understand the path female students take through STEM, educators and policymakers have called for research in mathematical self-efficacy and mathematical identity (Froschl & Sprung, 2016). In the present dissertation, the gender gap will be studied in all variables, but the longitudinal relationships among the three variables will be analyzed separately for male and female students to see if the relationship between the variables depends on gender.

Chapter 3 Methods

This section will describe the methods used in this study. It will start by exploring the participants and measures used. Next, it will explain how missing data, a common problem in longitudinal research, was addressed. Then it will describe the analytic procedures used, including how the cross-lagged panel models were developed and analyzed. Finally, it will restate the research questions and forward the hypotheses.

Participants

The High School Longitudinal Study of 2009 (HSLs: 09) was conducted by the National Center for Education Statistics (NCES) in response to a policy mandate to collect and disseminate statistics and other data related to education in the United States and the need for policy-relevant, nationally representative longitudinal samples of secondary school students. The NCES used stratified random sampling in the study. A total of 944 schools participated in the first stage of the study. The follow up included all schools that participated in the base-year data collection. However, five schools are not included in the number of eligible schools; four had closed, and one did not have any base-year students still enrolled.

Among questionnaire-capable 9th grade students ($n = 24,658$), 21,444 completed the student questionnaire (LoGerfo, Christopher & Flanagan, 2011) in the base year study. In the follow up, 20,594 students completed the student questionnaire (Ingels & Dalton, 2013). In the first follow-up, all base-year sample members—including nonparticipants—were followed and asked to complete the questionnaire (instruments were available to both in-school and out-of-school sample members). The initial target population of students was defined to include all 9th grade students who attended the study-eligible schools at the time of the study. The follow-up was conducted in the spring of 2012 when most of the sample members were in 11th grade

(Ingels & Dalton, 2013). At both stages, the only exclusionary criterion was the capability of completing the questionnaire or assessment due to language barriers or severe disabilities. Students with moderate disabilities with an individualized education plan that specified the accommodations of extra time, frequent breaks, or a separate area, for the district and state testing, were allowed access to those accommodations while completing the measures in this study (Ingels & Dalton, 2013). NCES developed a nationally representative sample for socioeconomic status (SES) and race/ethnicity (Ingels et al., 2013). Information on the students in the study is provided in Table 1, which was gathered from Ingels et al. (2011) and Ingels et al. (2013).

Table 1. *Demographic Information*

	Base Year (9 th Graders) n (%)	Follow-Up (11 th Graders) n (%)
Total Participants	21444	20594
Male Students	10887 (50.8%)	10384 (50.4%)
Female Students	10557 (49.2%)	10210 (49.6%)
Base Year School Type		
Public	17511 (81.7%)	17164 (83.3%)
Private	3933 (18.3%)	3430 (16.7%)
Race/Ethnicity		
American Indian	233 (1%)	187 (1%)
Asian/Pacific Islander	2144 (10%)	2129 (10.3%)
Black	2684 (12.5%)	2518 (12.2%)
Hispanic	3516 (16.4%)	3193 (15.5%)
White	12630 (58.9%)	12217 (59.3%)
More than one race	247 (1.2%)	350 (1.7%)

Measures

Gender

The student's gender data is a composite variable taken from the base year student questionnaire, parent questionnaire, and school-provided sampling roster. On each of these assessments, the student or parent was asked to indicate whether the student was male or female. At the base year, if the gender indicated by any of the three sources was inconsistent, a

researcher manually reviewed the sample member's first name and selected the gender they thought best matched the name. This was corrected at the follow-up, which is the dataset being used here. At that time, the variable was updated when missing with data obtained from the student's questionnaire, on which they were asked to indicate male or female (Ingels et al., 2011; Ingels et al., 2013). Given how the sex/gender variable was assessed in the dataset, students were forced to identify as male or female while completing the survey. Thus it will not allow for conclusions about students who are intersex, transgender, non-binary, or otherwise do not identify as male or female.

Mathematics Achievement

The High School Longitudinal Study of 2009, used for the present study, utilized a Mathematics Assessment in Algebraic Reasoning to measure mathematics achievement. Mathematical knowledge is often divided into two types: procedural knowledge and conceptual knowledge. One method used to measure mathematics achievement that includes the integration of procedural and conceptual knowledge is standardized tests (Knuth, Alibali, McNeil, Weinberg, & Stephens, 2005). Algebra is considered by many to be a gatekeeper in school mathematics, critical to further study in mathematics as well as to future educational and employment opportunities. Yet, many students have difficulty learning algebraic reasoning (National Research Council [NRC], 1998). Recent reform efforts in mathematics education have made algebra curricula and instruction a focal point (NRC, 1998; Ball, 2003). Standardized tests of algebraic reasoning, therefore, are considered essential assessments of math achievement for high schoolers (Knuth et al., 2005).

The HSLS-09 mathematics assessment was designed to provide a measure of the student achievement in algebraic reasoning at two points in time; in 9th grade and 11th grade. The test

framework was designed to assess the major content domains and algebraic processes (Ingels et al., 2011). The content domains measured included: The language of algebra; Proportional relationships and change; Linear equations, inequalities, and functions; Nonlinear equations, inequalities, and functions; Systems of equations; and Sequences and recursive relationships. The algebraic processes measured included: Demonstrating algebraic skills; Using representations of algebraic ideas; Performing algebraic reasoning; and Solving algebraic problems.

The HSLS-09 mathematics assessment score was derived from a two-stage computer administered test. In the first part, all students took a 15-item router test. The router test was made up of 25% low, 25% high, and 50% moderate complexity items, which were selected from a pool of items rated for difficulty. Based on this performance, each student was routed to a low, moderate, or high level of difficulty stage two test, which was designed for the student from a pool of items. Each stage two test, however, had some overlap of questions with the other tests in that stage of the same complexity level. All stage two tests consisted of 25 items. Students were allowed to skip and return to items within each stage. An online scientific calculator was available to them during all stages. At time one, the test was made up of 9th grade algebraic reasoning skill items, and at time two, the test was made up of 11th grade algebraic reasoning skill items. The skills needed were identified from the common core grade level standards.

Results from this assessment are available in several formats in the full dataset (Ingels et al., 2013). For this study, only the standardized theta score, as provided by the dataset, was used. This score was calculated from the item response theory (IRT)-based estimate of percent-correct. The standardized theta score provides a norm-referenced measurement of achievement that is an estimate of performance relative to the population as a whole. It includes information on status

compared to peers. At each time, the mean was separately standardized and rescaled to 50, with a standard deviation of 10 (Ingels et al., 2013). The reliability of the standardized theta score was found to be .92 at both the base year and follow up (Ingels et al., 2011; Ingels et al., 2013).

Mathematical Self-Efficacy

Items were written and selected for the mathematical self-efficacy scale by the National Center for Education Statistics with the intent to be used and analyzed as a scale of mathematical-self-efficacy (Ingels et al., 2011). The items were developed using a review of the literature, consultation with technical and methodological experts, and field testing to establish a theoretically valid measure of the construct (Ingels et al., 2011). The mathematical self-efficacy scale was constructed using responses to the following Likert scale items: "Confident s/he can do an excellent job on tests in current math class," "Understands most difficult textbook material in current math class," "Certain s/he can master skills being taught in current math class," and "Confident s/he can do an excellent job on current math class assignments" (Ingels et al., 2013). The Likert scale had four options, "strongly disagree," "disagree," "agree," and "strongly agree." The items within this set were recoded so that higher values indicated higher levels of mathematical self-efficacy. Cronbach's alpha estimate at time one was .90 (Ingels et al., 2011) and for time two was .89 (Ingels et al., 2013). In this study, the four individual items were indicators of a mathematical self-efficacy latent variable.

Mathematical Identity

Items were written and selected for the mathematical identity scale by the National Center for Education Statistics with the intent to be used and analyzed as a scale of mathematical identity (Ingels et al., 2011). The items were developed using a review of the literature, consultation with technical and methodological experts, and field testing to establish a

theoretically valid measure of the construct (Ingels et al., 2011). The mathematical identify scale was constructed using responses to the following Likert scale items: "Student sees her/himself as a math person" and "Others see the student as a math person" (Ingels et al., 2013). The Likert scale had four options, "strongly disagree," "disagree," "agree," and "strongly agree." The items within this set were recoded so that higher values indicated higher levels of mathematical identity. Cronbach's alpha estimate at time one was .84 (Ingels et al., 2011), and for time two was .88 (Ingels et al., 2013). In this study, the two individual items were used as indicators of a mathematical identity latent variable.

Missing Data

Longitudinal studies provide researchers a unique opportunity to examine the relationship between variables over time (Palmer & Royall, 2010). While this type of research is preferred for evaluating the nature of relationships between variables, longitudinal studies are not without challenges. One of the most significant obstacles to longitudinal studies is the presence of missing data due to attrition of study participants (Little, 2013; Young & Johnson, 2015). Attrition may result in selection bias, which can affect results from the statistical model (Little, 2013). For example, participants who completed the mathematics achievement test in the base-year, 2009, may be unique from students who completed the test in the follow-up year, 2011.

There are several methods available to handle and adjust for missing data. Historically, the default method for dealing with missing data was to exclude observations with missing data from the analysis using listwise or pairwise deletion methods (Palmer & Royall, 2010). Listwise deletion methods drop cases where there is missing data on any of the variables used in the analysis. In contrast, pairwise deletion techniques remove incomplete cases only from analysis that includes the variable with missing values (Roth, 1994). Unfortunately, there are several

disadvantages to relying on these traditional techniques, including a loss of power and increasing the probability of bias and error. These methods rely on only partial data, thereby increasing Type II error rates by reducing the total sample size. Moreover, unless the missing data can be assumed to be missing completely at random (and not due to any of the study variables), then these approaches will result in bias estimates of the true population (Palmer & Royall, 2011).

Currently, maximum likelihood (ML) estimation is considered the “gold standard” for missing data techniques (Baraldi & Enders, 2010; Schafer & Graham, 2002). The primary benefit to this method is that it does not delete cases or variables with missing data, nor does it attempt to use a single value to replace missing values, but instead uses all of the information provided to estimate the model parameters that best reflect the sample data (Baraldi & Enders, 2010; Little, 2013). The parameters of a statistical model are estimated in the presence of missing data by producing a different likelihood function for each different missing data pattern. These different likelihoods are then aggregated to compute the overall likelihood for a given set of model parameters, and a maximum likelihood is determined by which set of model parameters produces the largest overall likelihood of producing the sample data (Little, 2013). For this investigation, the ML estimation technique was used because the ML procedure is well suited for longitudinal analysis (Baraldi & Enders, 2010; Little, 2013), and it tends to be a more powerful technique than imputation-based methods (Graham, 2009).

Analytic Procedure

Measurement Model

Before the estimation of the structural model, a measurement model was developed with the six latent variables. Latent-variable structural equation modeling (SEM) often begins with a confirmatory factor analysis, or measurement model. This model is a crucial first step because it

tests the adequacy of the expected relations and constraints between the measured indicators and underlying latent variables (Little, 2013). The fit of the measurement model was evaluated to determine the adequacy of the measurement properties of each construct. Primary steps for the analysis included an evaluation of model fit and interpreting the parameter estimates. Next, the model was used for tests of factorial invariance.

Factorial Invariance

Any comparison of the same constructs across time or groups assumes that the measurements are factorially invariant (Little, 2013). In longitudinal research, age, maturation, experience, and personal choices are all factors that can influence change in the measures over time. Likewise, in research comparing groups, social, environmental, and personal choices are all factors that can affect change in a measure across groups. The critical question, therefore, is whether it was these influences that were what resulted in a change in the measured factor or if the change over time or across groups was a result of a measurement error within the items that make up the factor. For longitudinal models, the issue of invariance applies to the indicators and constructs that are measured at more than one occasion. Likewise, for across group models, invariance tests apply to factors that are measured in more than one group (Little, 2013).

There are different levels of testing for factorial invariance, which are configural, weak, strong, and strict. The first level, configural, simply requires that the relations between each indicator have been set in the same pattern of fixed and freed. That is, everywhere there is an estimate at one time point or in one group, there is a consistent estimate in the other time points or groups. Then the question is, does the pattern look the same, and does the model appear to fit reasonably well? The configurally invariant model is primarily used as the baseline model to

evaluate the degree to which the different levels of factorial invariance are supported by the data (Little, 2013).

The second level is weak factorial (i.e., factor loading) invariance. When testing for the weak level, the loadings of the indicators on the constructs are estimated as a common set of parameters. That is, they are not allowed to vary across groups or times. This is done by setting the estimates to be equal across measurement occasions. The fit of this model can then be compared to a baseline model. If the fit of the model is about the same, then this suggests that there is a consistent relative relationship, which suggests invariance (Little, 2013).

Longitudinal Factorial Invariance

Longitudinal factorial invariance tests the assumption that the same constructs across time are fundamentally similar. When observed cross-time construct differences exist, yet the constructs are factorially invariant, those differences are likely due to the factors, not to fundamental measurement differences (Little, 2013). Factorial invariance requires that increasingly restrictive constraints be placed on model parameters and tested in sequence (Keith & Reynolds, 2012; Meredith, 1993).

Following the construction of the measurement model, longitudinal factorial invariance was tested to address the assumption that the same constructs across time were fundamentally the same. To do so, the corresponding factor loadings across like latent variables were fixed to be equal, and the fit of the model was compared to the model without the fixed corresponding loadings. When observed cross-time construct differences exist, yet the constructs are factorially invariant, those differences are likely due to differences in the constructs, not to fundamental measurement differences (Little, 2013).

Multiple Group Factorial Invariance

As with longitudinal data, measurement invariance constraints provide a stable baseline for making meaningful construct comparisons across groups. When invariance is found, it assures the researcher that the measurement is the same across groups, and that suggests that there is no evidence of measurement anomalies. Following the construction of the measurement model, multiple group factorial invariance was tested across gender to address the assumption that the same constructs across groups were fundamentally the same. To do so, the corresponding factor loadings across like latent variables were fixed to be equal, and the fit of the model was compared to the configural model. When observed cross-group construct differences exist, yet the constructs are factorially invariant, those differences are likely due to differences related to the constructs, not to fundamental measurement differences (Little, 2013).

Structural Model

Cross-lagged panel analysis, an (SEM) technique, was used to test the study's hypotheses. As a way of ameliorating measurement error—a common threat to validity in causal analysis—a latent variable approach was used (Shadish, Cook, & Campbell, 2002). A single factor was developed for each self-belief construct, mathematical self-efficacy, and mathematical identity. For math achievement, a single indicator latent variable was created by fixing the residual variance to the reliability estimate. All analyses were done in SPSS and AMOS.

The best method to assess the relationship between mathematical self-efficacy, mathematical identity and mathematics achievement has remained a challenge (Talsma, Schüz, Schwarzer, & Norris, 2018). This is, in part, because mathematical self-efficacy, mathematical identity, and past mathematics achievement are expected to be covarying common-cause variables (Bandura, 2012; Middleton, Mangu, & Lee, 2019; Talsma, Schüz, Schwarzer, &

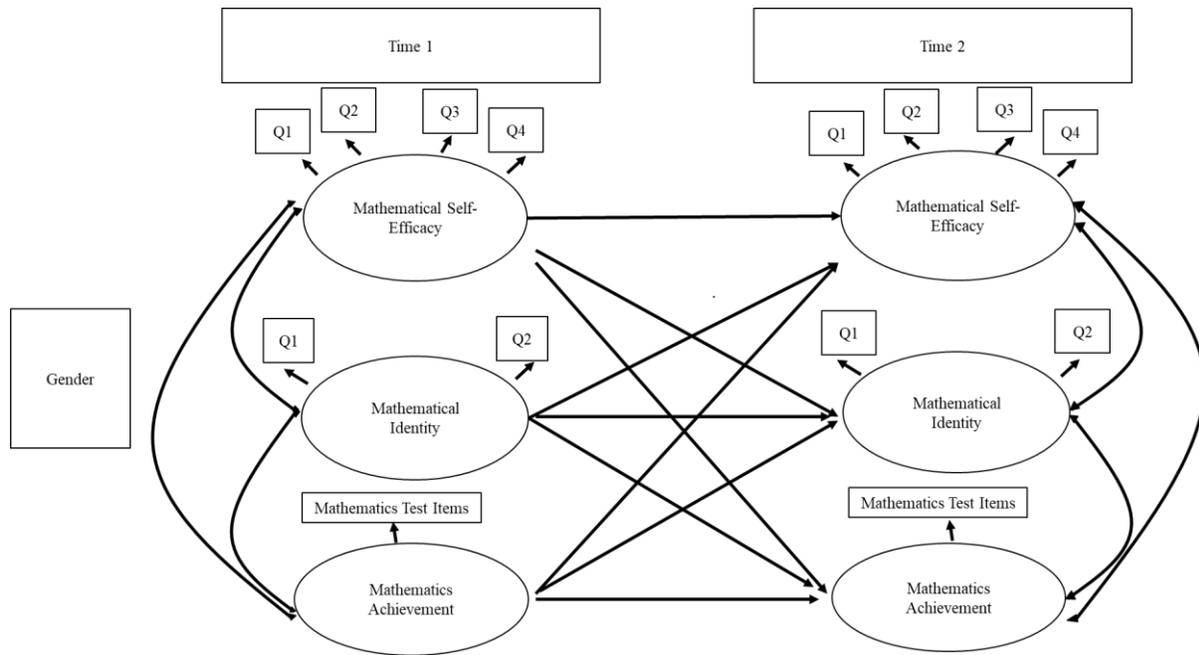
Norris, 2018). As a result, any approach that does not allow for control will result in inflated estimates of the influence of these variables on each other. Therefore, a procedure is required where mathematical self-efficacy, identity, and achievement performance can be controlled at time one (Little et al., 2007).

Cross-lagged panel models of longitudinal, reciprocal effects are well suited to investigate the nature of self-belief and achievement change over time because they allow for an exploration of the direction of the effects of self-beliefs and achievement while controlling for previous measures of these same constructs (Little et al., 2007). Panel data consists of at least two variables measured at two or more time-points in the same set of subjects. Analysis of panel data has been recognized for its advantages in testing for causal effects because it can provide evidence regarding all three conditions of causality (Finkel, 1995). Panel models present a remarkably powerful approach to chicken-and-egg questions (Tyagi & Singh, 2014). It provides the means to address the question: “Which is the more important influence, self-efficacy on achievement, or achievement on self-efficacy?” (Greenberg, 2008).

The analysis included mathematical self-efficacy, mathematical identity, and mathematics achievement latent variables from when the students are in 9th grade (time one) and 11th grade (time two). The model can be seen in Figure 1. The observed variables (boxes in the model) are indicators of the latent or unobserved variables (circles in the model). Unidirectional arrows from the latent variables to the indicators represent factor loadings. The structural model specifies the hypothesized pattern of causal influence between the latent variables. Two-headed arrows represent the hypothesized covariance or correlations. Time two two-headed arrows represent the covariance (correlations) among the residual variances of the latent variables. After estimating a model with the whole sample, two models were constructed that match Figure 1,

one model was for male students, and the other was for female students, in a multiple-group model. Then, the relationships between constructs were compared to see if there was a difference in the size and nature of the relationships for female and male students. Finally, the constructs were progressively fixed to be equal to test if gender was a moderating variable.

Figure 1: Latent variable cross-lagged panel model of the relationship of mathematical self-efficacy, mathematical identity, and mathematics achievement



Model Evaluation

Standalone fit indexes were used in this study to examine model fit to ensure that the structural models were of adequate fit to allow for comparing paths. There are a variety of commonly used indicators of model fit. Several commonly found indicators were used in this analysis; those indicators were the root mean square error of approximation (RMSEA), comparative fit index (CFI), and the Chi-square test (χ^2).

The χ^2 test is one of the most common measures of model fit (Keith, 2006). The χ^2 goodness of fit tests the difference between the observed data and the hypothesized model. When testing for model fit, a smaller χ^2 value is desired because a smaller value indicates that the

hypothesized model fits the data. For example, $\chi^2 = 0$ represents a perfect fit of the model-implied covariance matrix to the observed covariance matrix. It is important to note that the sample size influences χ^2 . Specifically, in models that have more than 400 cases, the χ^2 statistic is almost always statistically significant (Kenny, 2015), which could lead to a Type II error. To further assess model fit and reduce the probability of committing a Type II error, the CFI and RMSEA fit indices were included.

The RMSEA provides an index of the amount of misfit per degrees of freedom in the model. An RMSEA that is .08 or lower is considered to suggest acceptable model fit, with .05 or lower suggesting good fit. The CFI is the ratio of misfit of the tested model to the misfit of the null model. A CFI above .90 is considered to indicate acceptable fit, with a CFI above .95 suggesting a good fit (Little, 2013).

Standalone and comparative fit statistics were used when examining models for longitudinal and multiple group invariance. When evaluating for longitudinal and multiple group invariance of the factor loadings, χ^2 tests are highly influenced by population size. They are considered to be too sensitive to slight fluctuations in the context of invariance testing (Little, 2013). As the sample size for this study is large, an alternative way to evaluate invariance is to compare the size of a CFI change. Cheung and Rensvold (2002) provided now widely used guidelines suggesting that if the change in CFI is less than .01, then the assumption of factorial invariance is reasonable. The χ^2 test, however, in addition to the CFI, was used to evaluate the change in fit when structural parameters were fixed equal across gender.

Research Questions

The following research questions and hypotheses guided this study:

Question 1. What is the temporal relationship between mathematical self-efficacy, mathematical identity, and mathematics achievement?

Hypothesis 1: There will be a significant temporal relationship among each of the key variables demonstrated by reciprocal longitudinal effects between mathematical self-efficacy, mathematical identity, and mathematics achievement. The nature of the temporal relationship is expected to be positive between mathematical self-efficacy and mathematics achievement, and between mathematical self-efficacy and mathematical identity. The relationship of mathematical identity with mathematical self-efficacy may result in a weaker relationship between mathematical identity and mathematics achievement.

Question 2: Do the observed temporal relationships among mathematical self-efficacy, mathematical identity, and mathematics achievement depend on gender?

Hypothesis 2: Although female students are predicted to show the same mathematics achievement and lower mathematical identity and self-efficacy than male students, based on prior research, the temporal relation between mathematical self-efficacy and mathematics achievement are predicted not to differ between male and female students. If gender does moderate the relationships, it is predicted that self-beliefs will be affected more strongly by achievement in male than in female students.

Chapter 4: Results

Two main research questions guided this study. Research question 1 asked about the temporal relationship between mathematical identity, mathematical self-efficacy, and mathematics achievement. Research question 2 asked about a potential moderating role that student gender has on the nature of the temporal relationship between the variables. This chapter reports descriptive statistics, the results from the cross-lagged panel model used to answer research question 1, and from the multi-group cross-lagged panel model to answer question 2.

Item Level Descriptive Statistics

The data were screened relative to the assumptions of SEM procedures (Kline, 2005) using SPSS 26 statistical software. Each variable's distribution was evaluated for normality by examining histograms and statistics for skewness and kurtosis. All variables consistently approximated a normal distribution. One item associated with the mathematics identity latent variable at the follow up, "Teenager sees himself/herself as a math person," showed a kurtosis of -1.05, suggesting that this distribution had slightly lighter tails and a flatter peak than a normal distribution. Other than that variable, no other item had skewness or kurtosis value exceeding the absolute value of one. Each variable is within the range of a normal distribution (Mertler & Vannatta, 2005), and skewness and kurtosis values were well below the minimums that have been recommended for the use of maximum likelihood estimation. This assessment concluded that deviations from normality were insubstantial so that standard SEM analysis and fit indices were used. Kurtosis, skewness, and bivariate correlations across all variables at both time-points can also be found in Table 2. The sample size, means, and standard deviations for the total sample and also for male and female students can be found in Table 3.

Table 2. Bivariate Correlations, Skewness & Kurtosis for Observed Variables

Measure	9 th Grade Math Score (IMTS)	9 th Grade sees self as math person (MPer1)	9 th Grade others see as math person (MPer2)	9 th Grade can do excellent on math tests (1MTest)	9 th Grade can understand math text (1Mtext)	9 th Grade can master math skills (1MSkill)	9 th Grade can do excellent on math assignment (1MAssi)	11 th Grade Math Score (2MTS)	11 th Grade sees self as math person (2MPer1)	11 th Grade others see as math person (2MPer2)	11 th Grade can do excellent on math tests (2MTest)	11 th Grade can understand math text (2MText)	11 th Grade can master math skills (2MSkill)	11 th Grade can do excellent on math assignment (2MAssi)
IMTS	-													
1MPer1	.35	-												
1MPer2	.38	.73	-											
1MTest	.27	.51	.46	-										
1MText	.25	.49	.44	.68	-									
1MSkill	.29	.49	.45	.68	.67	-								
1MAssi	.28	.48	.44	.73	.64	.72	-							
2MTS	.75	.34	.38	.27	.26	.29	.27	-						
2MPer1	.34	.55	.46	.36	.34	.33	.32	.37	-					
2MPer2	.37	.49	.49	.34	.33	.32	.31	.40	.79	-				
2MTest	.22	.30	.27	.30	.28	.28	.27	.25	.51	.48	-			
2MText	.20	.31	.27	.28	.31	.28	.26	.23	.51	.48	.65	-		
2MSkill	.25	.29	.28	.28	.27	.30	.27	.29	.49	.46	.74	.63	-	
2MAssi	.23	.29	.27	.30	.27	.29	.29	.27	.48	.46	.76	.61	.27	-
Sex	.004*	-.08	-.02	-.11	-.11	-.07	-.06	-.009*	-.10	-.07	-.12	-.11	-.08	-.08
Skew	-.06	-.10	-.14	-.46	-.22	-.47	-.57	.06	.09	-.02	-.31	-.13	-.44	-.57
Kurtosis	-.13	-.91	-.79	-.01	-.46	.20	.43	-.31	-1.05	-.87	-.37	-.71	-.04	.43

Note. IMTS = Mathematics Standardized Theta Score; MPer1 = Student sees himself/herself as a math person; MPer2 = Others see student as a math person; MTest = Student confident can do excellent job on math tests; MText = Student certain can understand math textbook; MSkill = Student certain can master skills taught in math course; MAssi = Student confident can do excellent job on math assignments
 Negative numbers for sex indicate lower scores for female students
 *Correlation was not statistically significant. All other correlations were statistically significant $p < .01$

Table 3. *Sample Size, Means, & Standard Deviations for Observed Variables in Total Sample and Male and Female Students*

Variables & Composites	Total Sample <i>N</i>	Total Sample Mean (<i>SD</i>)	Female Student <i>N</i>	Female Student Mean (<i>SD</i>)	Male Student <i>N</i>	Male Student Mean (<i>SD</i>)
9 th Grade Math Score (1MTS)	21444	51.11 (10.08)	10557	51.15 (9.66)	10887	51.07 (10.47)
9 th Grade Math Identity Composite	21159	.04 (1.00)	10457	-.01 (1.00)	10702	.09 (1.00)
9 th Grade sees self as math person (MPer1)	21347	2.51 (.95)	10825	2.44 (.94)	10522	2.59 (.95)
9 th Grade others see as math person (MPer2)	21183	2.53 (.91)	10463	2.51 (.90)	10720	2.54 (.92)
9 th Grade Math Self-Efficacy Composite	18759	.04 (1.00)	9326	-.05 (.99)	9433	.14 (.99)
9 th Grade can do excellent on math tests (1MTest)	19049	2.98 (.76)	9468	2.90 (.77)	9581	3.06 (.74)
9 th Grade can understand math text (1Mtext)	19006	2.72 (.82)	9452	2.63 (.82)	9554	2.81 (.81)
9 th Grade can master math skills (1MSkill)	18976	2.98 (.73)	9429	2.93 (.72)	9547	3.03 (.73)
9 th Grade can do excellent on math assignment (1MAssi)	18926	3.07 (.72)	9412	3.03 (.72)	9514	3.11 (.72)
11 th Grade Math Score (1MTS)	20594	51.50 (10.15)	10210	51.41 (10.62)	10384	51.59 (10.62)
11 th Grade Math Identity Composite	20024	.053 (1.02)	9934	-.04 (1.02)	10090	.15 (1.00)
11 th Grade sees self as math person (MPer1)	20103	2.40 (1.00)	9968	2.29 (1.00)	10135	2.50 (.98)
11 th Grade others see as math person (MPer2)	20048	2.45 (.93)	9948	2.38 (.93)	10100	2.51 (.92)
11 th Grade Math Self-Efficacy Composite	19771	.04 (1.00)	9840	-.07 (1.01)	9931	.15 (.98)
11 th Grade can do excellent on math tests (1MTest)	19998	2.78 (.82)	9954	2.67 (.83)	10044	2.88 (.80)
11 th Grade can understand math text (1Mtext)	20018	2.55 (.89)	9961	2.45 (.89)	10057	2.65 (.88)
11 th Grade can master math skills (1MSkill)	20017	2.85 (.78)	9957	2.79 (.78)	10060	2.91 (.77)
11 th Grade can do excellent on math assignment (1MAssi)	19983	2.92 (.77)	9944	2.85 (.78)	10039	2.98 (.75)

Note. 1MTS =Mathematics Standardized Theta Score; MPer1 =Student sees himself/herself as a math person; MPer2 =Others see student as a math person; MTest =Student confident can do excellent job on math tests; MText =Student certain can understand math textbook; MSkill =Student certain can master skills taught in math course; MAssi=Student confident can do excellent job on math assignments

NCES created the scales for the composite variables of mathematical self-efficacy and mathematical identity. When creating the scales, the variables were standardized with a mean of 0 and a standard deviation of 1 (Ingels et al., 2011; Ingels et al., 2013). The items that made up the scales were measured using a Likert scale with four options, “strongly disagree,” “disagree,”

“agree,” and “strongly agree.” While male and female students’ mathematics achievement, as measured by the standardized theta score was approximately equal, female students did show lower self-ratings on the mathematical self-efficacy and mathematical identity scales as well as every mathematical self-efficacy and mathematical identity item in both 9th and 11th grade. All bivariate correlations were statistically significant at the $p < .01$ level except for the relationship between the student’s gender and the mathematics standardized theta score, which was not statistically significant.

Missing Data

Before conducting preliminary and primary analyses, an analysis of missing data was completed to determine the amount of missing data present in the current sample. Missing data was expected because the High School Longitudinal Study of 2009 was a national study involving tens of thousands of participants. Therefore, some attrition is expected. No variables contained more than 20% missing data. There were 15518 complete cases out of a total sample of 23503 students. Little’s test for MCAR (Little, 2013) was statistically significant for the entire sample $\chi^2(982) = 1431.60, p < .001$, suggesting that the data cannot be assumed to be MCAR. Case deletion assumes MCAR. In this study, maximum likelihood estimation was used, which assumes missing data are MAR.

Before making the data available to the public, the HSLs:09 assessment data were examined for possible indicators of a lack of motivation to answer questions to the best of the student’s ability. Examples of potential signs are missing responses and pattern marking (e.g., all answers were “A” or “ABCDABCDABCD...”). About 1% of the test records from the base year, and follow-up was discarded from the analysis sample for either a high percentage of missing responses or for showing a pattern (Ingels et al., 2011; Ingels et al., 2013). After the

elimination of these cases, no particular reasons were forwarded for why the data may not demonstrate MCAR. However, the researchers did note that “HLSL:09 variables, in general, did not suffer from high levels of item nonresponse” (Ingels et al., 2011, p. 162).

It is not possible to directly test for MAR (Enders, 2010). However, there is no known evidence to suggest that the reason for missing data was directly caused by the variables themselves (mathematical self-efficacy, mathematical identity, or mathematics achievement), which would be the case for MNAR. Preliminary and primary analyses were conducted using all data, and data were assumed to be MAR.

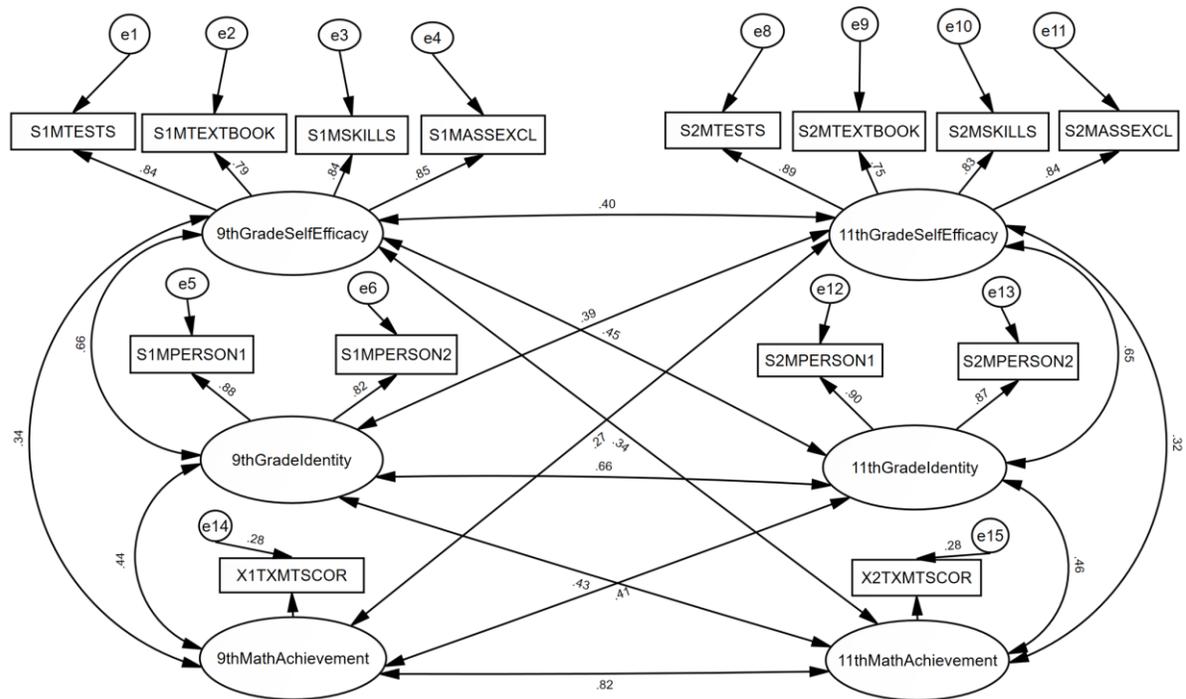
Measurement Model

Before testing the causal model, the measurement model was examined through confirmatory factor analysis (CFA) to increase confidence that the specified indicators for mathematical self-efficacy and mathematical identity were indeed capturing separate constructs in the sample data. First, items or test scores were assigned to their respective latent factors of mathematical self-efficacy, mathematical identity, and mathematics achievement, and with the latent factors being allowed to correlate freely, as is illustrated in Figure 2. Next, longitudinal invariance was tested by fixing corresponding factor loadings across time to be equal.

In the Longitudinal Measurement Model for the total sample, factor loadings were consistently strong (range of standardized loadings for mathematical self-efficacy indicators was .75 to .89, and for mathematical identity, indicators was .82 to .90). The correlations ($r = .66$ and $r = .65$) between the mathematical identity and self-efficacy latent factors suggest that they are correlated, yet independent constructs. Model indices suggested excellent fit: RMSEA = .040, and CFI = .986. See the Longitudinal Measurement Model in Table 4 for fit statistics.

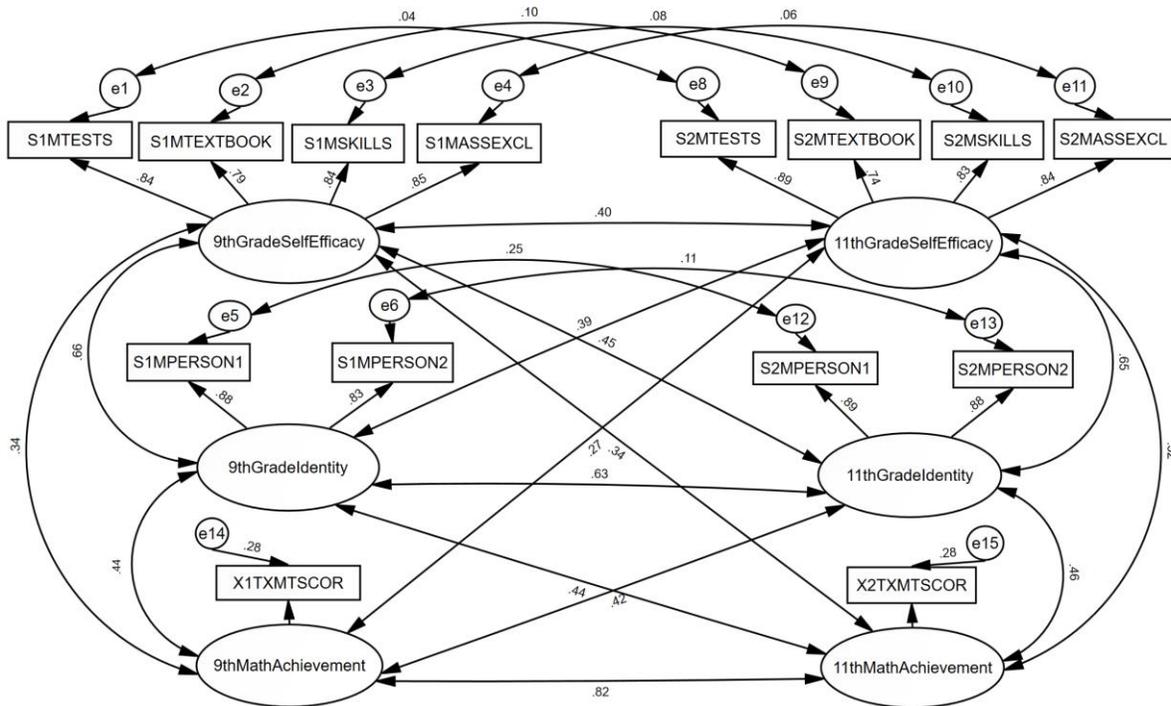
Despite the good fit, an additional model, as depicted in Figure 3, was tested (i.e., Correlated Residual Longitudinal Measurement Model). In this model, the time one item residual variance was correlated with its corresponding time two item residual variance. The reason for allowing these corresponding residuals to correlate across time is that the items have identical wording, so they may share common variance due to the exact wording, and this variance is not accounted for by their respective common factor. The factor loadings remained consistently strong in this model (range of standardized loadings for mathematical self-efficacy indicators .74 to .89; for mathematical identity indicators, .83 to .88), and model indices suggested excellent fit: RMSEA = .034, and CFI = .991. Model fit improved according to change in χ^2 and the other indices showed minimally improved fit (see Table 4). The correlations between the residuals were small in size, but statistically significant. Because of the improvement in fit, all models going forward included these residual correlations.

Figure 2. Longitudinal Measurement Model for the Total Sample



Before moving forward with substantive tests, it was important to test whether the constructs being measured in this study were the same across the two time points. This test involved constraining corresponding factor loadings to be equal across time one and time two for the mathematical identity and mathematical self-efficacy factors. The Correlated Residual Longitudinal Measurement Model is the “configural invariance model” (see Table 4 and Figure 3). This model already showed excellent fit to the data suggesting the model is an accurate representation of the observed data. If the Factor Loading Invariance Model does not result in a substantial degradation in model fit when compared to the Correlated Residual Longitudinal Measurement model, it suggests that the constructs are the same across time one and time two. For example, mathematical identity at time one is the same as mathematical identity at time two. If this model is acceptable, the constraints on those factors will be carried forward to subsequent structural models used to answer research question 1.

Figure 3. *Correlated Residual Longitudinal Measurement Model*



The result from the longitudinal invariance test is shown in Table 4 below. Invariance of the factor loadings was supported. While there was a statistically significant change in fit according to change in χ^2 , there was not a change in the CFI or RMSEA with the additional constraints, suggesting that the assumption of factorial invariance is reasonable (Cheung & Rensvold, 2002; Little, 2013). The large sample size is likely the cause of the statistically significant change in χ^2 (Little, 2013). The constructs were the same across time.

Table 4. *Model Fit Statistics for Longitudinal Measurement Models with the Total Sample*

Model Tested	χ^2 (df)	$\Delta \chi^2$ (Δ df)	<i>p</i>	RMSEA	CFI
Longitudinal Measurement Model	2470.93 (64)	-	<.001	.040	.986
Correlated Residual Longitudinal Measurement Model ^a	1629.48 (58)	841.45 (6)	<.001	.034	.991
Factor Loading Invariance Measurement Model ^{ab}	1723.56 (62)	94.08 (4)	<.001	.034	.991

Note. ^aFit is compared with the previously listed model.

^bThe fit of this model is identical to the fit of the Longitudinal Structural Model used to answer research question 1 and shown in Figure 4. They are equivalent models with regard to fit.

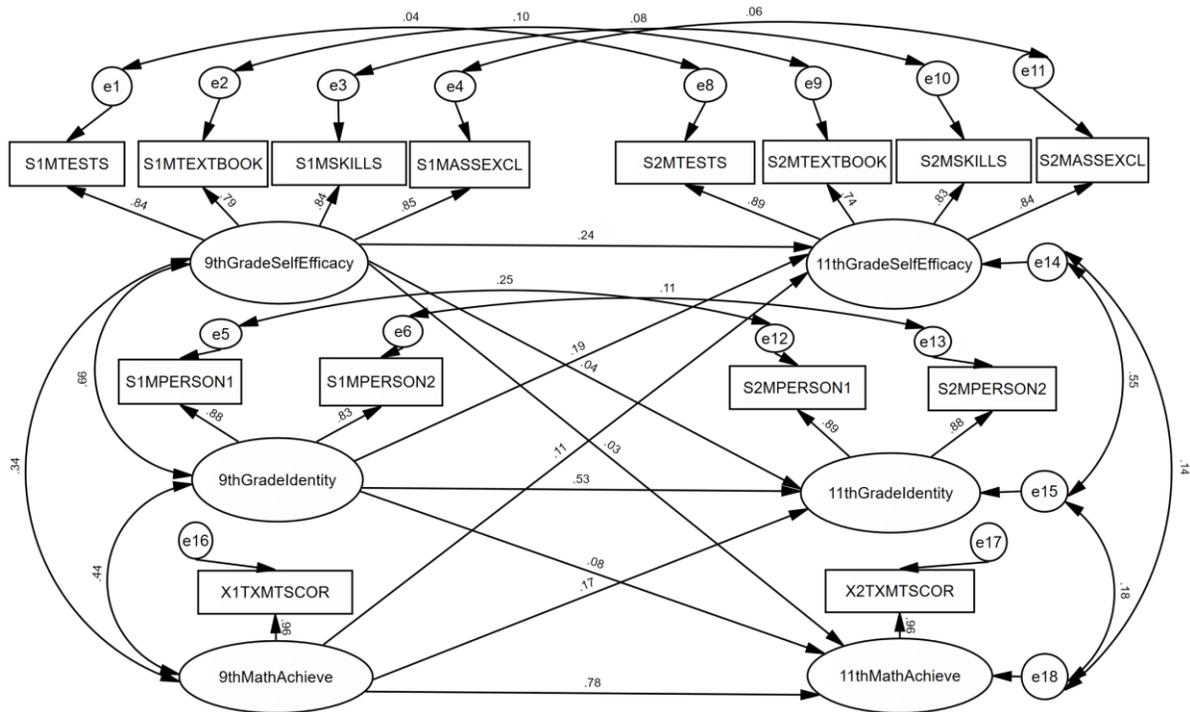
Research Question 1

The first research question asked about the temporal relationships between mathematical self-efficacy, mathematical identity, and mathematics achievement. After developing an appropriate measurement model, a structural model was estimated to investigate the temporal relations within and between mathematical self-efficacy, mathematical identity, and mathematics achievement (see Figure 4). The difference between the measurement model and the structural model is that in the structural model, the correlations between time one and time two factors were changed from correlations to directed paths that can be interpreted as regression weights. The fit the Longitudinal Structural model is the same as the Factor Loading Invariance Measurement Model shown in Table 4.

Unstandardized and standardized regression weight estimates and results of the statistical significance tests within constructs over time are reported in Table 5. Mathematics achievement

was mostly stable across the two-year period ($\beta = .776, b = .786, p < .001$). While less stable than mathematics achievement, mathematical identity also remained fairly stable ($\beta = .526, b = .566, p < .001$). In contrast, mathematical self-efficacy was relatively unstable from 9th to 11th grade ($\beta = .236, b = .268, p < .001$).

Figure 4. Total Sample Longitudinal Structural Model



In addition, unstandardized and standardized regression weight estimates and results of the statistical significance tests between constructs over time are reported in Table 5. When examining Pearson correlations, a correlation of .1, .2, and .3 are considered a small, medium, and large effect size, respectively (Gignac & Szodorai, 2016). The strongest temporal relationships between constructs were between 9th grade identity and 11th grade self-efficacy ($\beta = .189, b = .165, p < .001$) and between 9th grade achievement and 11th grade identity ($\beta = .169, b = .016, p < .001$). These relationships show medium effect sizes and are typical of those found in individual differences in research (Gignac & Szodorai, 2016). Small effect sizes were found

between 9th grade achievement and 11th grade self-efficacy ($\beta = .109, b = .008, p < .001$) and between 9th grade identity and 11th grade achievement ($\beta = .079, b = .936, p < .001$) (Gignac & Szodorai, 2016). There were two instances of statistical significance, due to the large sample size, in the presence of an unmeaningful paths. Specifically, the path coefficients for 9th grade mathematical self-efficacy to 11th grade mathematics achievement ($\beta = .025, b = .386, p = .002$) and 9th grade self-efficacy to 11th grade identity ($\beta = .045, b = .063, p < .001$) were too small to interpret meaningfully despite being statistically significant (Gignac & Szodorai, 2016).

Table 5. *Unstandardized and Standardized Regression Weight Estimates of Temporal Relations Within and Between Constructs*

Measure	UnStd. (S.E.)	Std.	p
Within Constructs			
9th Grade self-efficacy → 11th Grade self-efficacy	.268 (.013)	.236	<.001
9th Grade identity → 11th Grade identity	.566 (.012)	.526	<.001
9th Grade achievement → 11th Grade achievement	.786 (.006)	.776	<.001
Between Constructs			
9th Grade self-efficacy → 11th Grade identity	.063 (.015)	.045	<.001
9th Grade self-efficacy → 11th Grade achievement	.386 (.124)	.025	.002
9th Grade identity → 11th Grade self-efficacy	.165 (.011)	.189	<.001
9th Grade identity → 11th Grade achievement	.936 (.100)	.079	<.001
9th Grade achievement → 11th Grade self-efficacy	.008 (.001)	.109	<.001
9th Grade achievement → 11th Grade identity	.016 (.001)	.169	<.001

Note. UnStd.=Unstandardized regression coefficient; Std.=Standardized regression coefficient, S.E.=Standard Error

The hypothesis, therefore, that there would be a significant temporal relationship among each of the key variables demonstrated by reciprocal longitudinal effects between mathematical self-efficacy, mathematical identity, and mathematics achievement was partially supported.

While there were significant paths found in all relationships, in two instances, these paths were statistically significant due to sample size and were too small to interpret or suggest a causal relationship (Gignac & Szodorai, 2016). These findings indicate that in the total sample, 9th

grade mathematical self-efficacy had no causal influence on change in mathematical identity in 11th grade or mathematics achievement. All other relationships were found to be statistically significant and meaningful.

Research Question 2

The second research question asked whether the temporal relationships between mathematical self-efficacy, mathematical identity, and mathematics achievement were the same for female and male students. To test this hypothesis, the data were split into two groups by gender. Model fit was first tested separately by gender, and results are shown in Table 6.

Table 6. *Model Fit Statistics for Male and Female Longitudinal Measurement and Structural Models*

Model Tested	χ^2 (df)	$\Delta \chi^2$ (Δ df)	p	RMSEA	CFI
Female Students					
Longitudinal Measurement Model	1405.24 (64)	-	<.001	.043	.985
Correlated Residual Longitudinal Measurement Model ^a	977.51 (58)	427.73 (6)	<.001	.037	.989
Factor Loading Invariance Measurement Model ^a	1011.80 (62)	34.30 (4)	<.001	.036	.989
Male Students					
Longitudinal Measurement Model	1075.06 (64)	-	<.001	.036	.988
Correlated Residual Longitudinal Measurement Model ^a	686.19 (58)	388.87 (6)	<.001	.030	.993
Factor Loading Invariance Measurement Model ^a	754.29 (62)	68.10 (4)	<.001	.031	.992

Note. ^aFit is compared with the previously listed model.

Measurement Models by Gender

The Longitudinal Measurement Models for male and female students are displayed in Figures 5 and 6, respectively. The Longitudinal Measurement Model was set up with latent factors of mathematical self-efficacy, mathematical identity, and mathematics achievement, each being assigned the respective indicators, and with the latent factors being allowed to correlate freely. These models were the same as the Longitudinal Measurement Model in the total sample, except that here the model was run separately by gender. All model indices showed good fit, fit statistics are reported in Table 6.

Figure 5. Longitudinal Measurement Model for the Female Student Sample

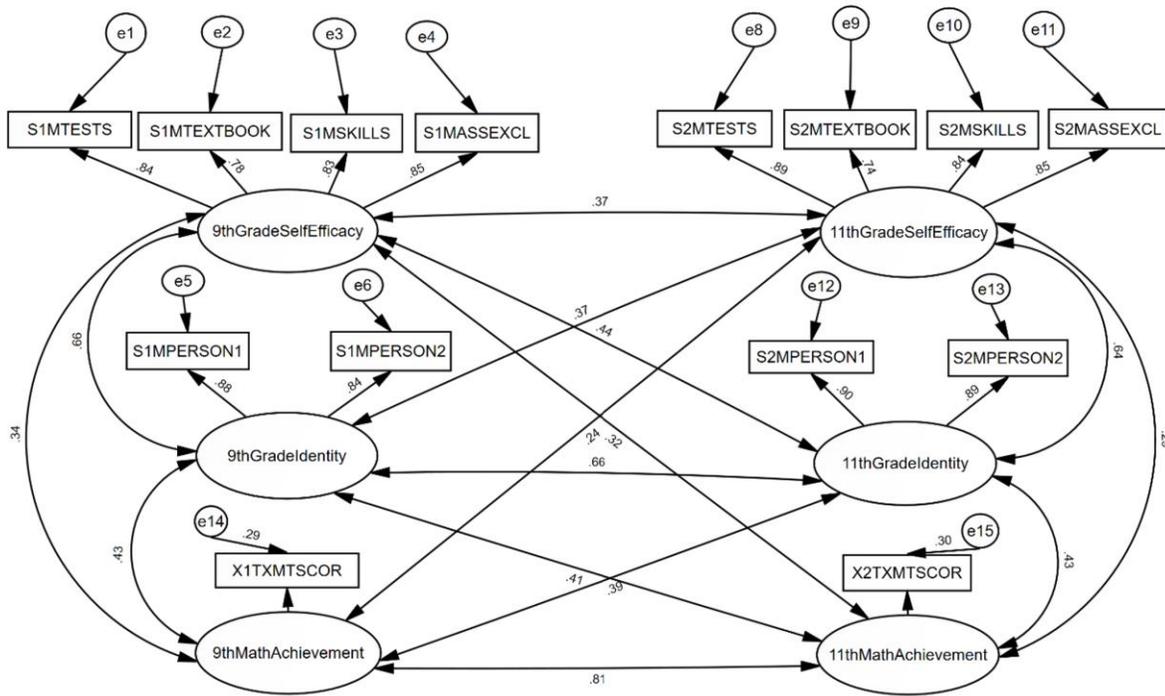
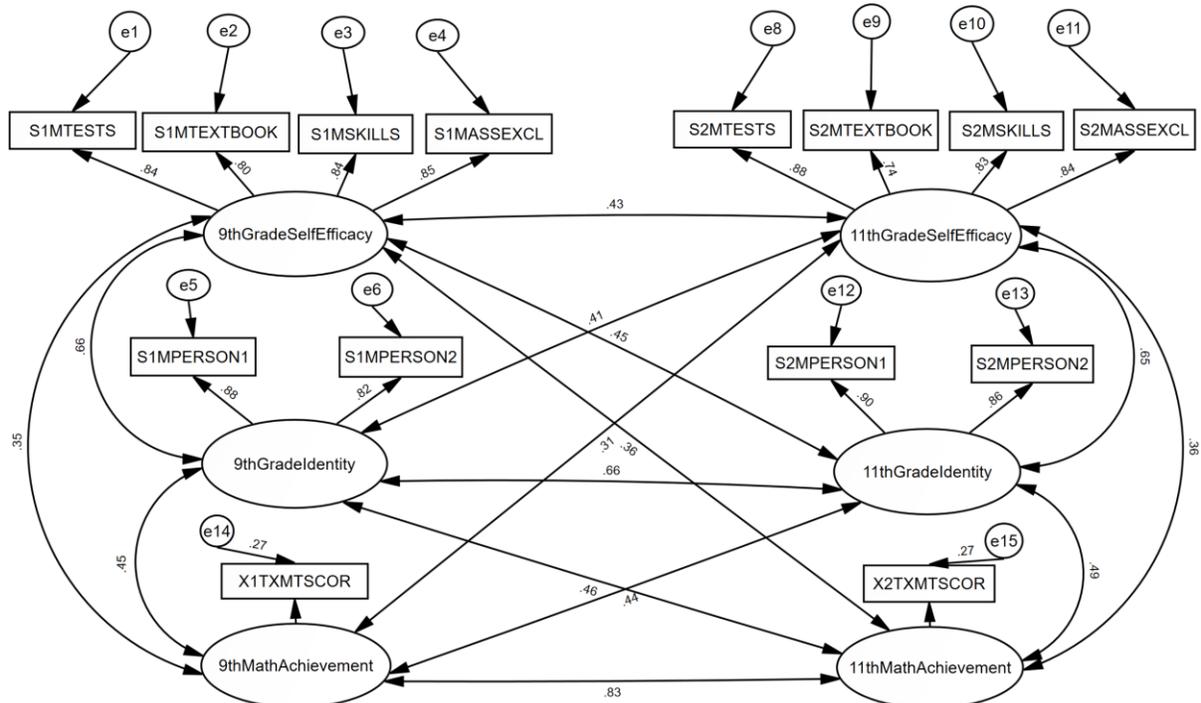


Figure 6: Longitudinal Measurement Model for the Male Student Sample



Despite the good fit, as with the total sample, the Correlated Residual Longitudinal Measurement Model was also tested for the male and female student samples. In these models, the time one item residual variance was correlated with its time two item residual variance. As with the total sample, the reason for allowing these corresponding residuals to correlate across time is that the items have identical wording, so they may share common variance due to the exact wording, and this variance is not accounted for by their respective common factor. These models are displayed in Figure 7 and Figure 8. All model indices suggested good fit for the male and female Correlated Residual Longitudinal Measurement Models (see Table 6). Model fit improved according to change in χ^2 . Therefore, all models going forward included these residual correlations.

Before moving forward with substantive tests for research question two, it was important to test whether the constructs being measured in this study were the same across the two time points in male and female samples. Similar to the total sample, this test involved constraining corresponding factor loadings to be equal across time one and time two for the mathematical identity and mathematical self-efficacy factors.

Results from longitudinal invariance tests are shown in Table 6 above and suggest factor loadings were invariant. While there were statistically significant changes in χ^2 , CFI and RMSEA changes were less than .01 with the additional constraints, and this suggests that the assumption of factorial invariance is reasonable (Cheung & Rensvold, 2002; Little, 2013). The large sample size is likely the cause of the statistically significant change in χ^2 (Little, 2013).

Figure 7. Female Correlated Residual Longitudinal Measurement Model

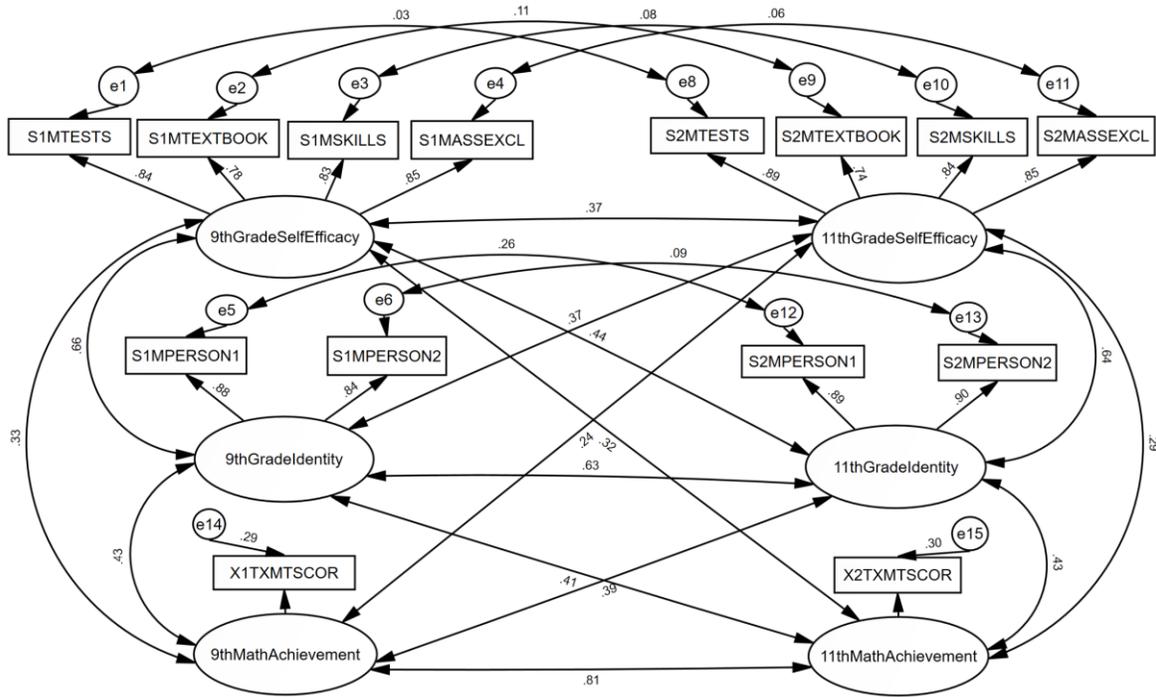
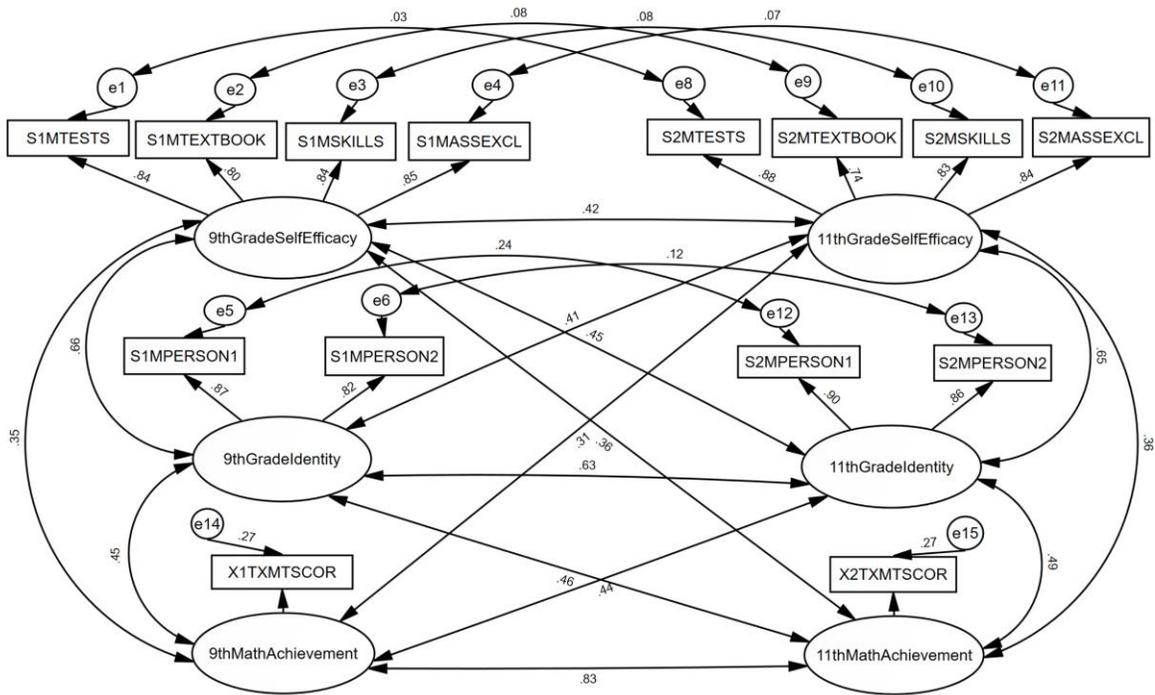


Figure 8: Male Correlated Residual Longitudinal Measurement Model



Multi-Group Invariance Testing

After confirming longitudinal invariance for male and female students, multiple group factorial invariance was tested to address the assumption that the constructs across gender were fundamentally the same. To do so, a configural model was developed, then the corresponding factor loadings across like latent variables were fixed to be equal, and the fit of the model was compared to the configural model. The Configural Invariance Model included the longitudinal constraints and essentially combined the fit for the two Factor Loading Invariance Measurement models shown in Table 7. Model fit for the configural and factorial invariance models across gender are also displayed in Table 7.

Table 7. *Model Fit Statistics for Configural and Factorial Invariance and for Tests of Moderation Across Gender*

Model Tested	χ^2 (df)	$\Delta \chi^2$ (Δ df)	<i>p</i>	RMSEA	CFI
Configural Invariance Model	1766.09 (124)		<.001	.024	.991
Factorial Invariance Across Gender Model	1797.28 (128)	31.19 (4)	<.001	.024	.990
Fixed Structural Path Model Achievement → Self-Efficacy ^a	1806.02 (1)	8.73 (1)	.003	.024	.990
Fixed Structural Path Model Achievement → Identity ^a	1801.67 (1)	4.39 (1)	.036	.023	.990
Fixed Structural Path Model Identity → Identity ^a	1804.59 (1)	7.30 (1)	.007	.024	.990
Fixed Structural Path Model Self-Efficacy → Identity ^a	1798.12 (1)	0.84 (1)	.359	.023	.990
Fixed Structural Path Model Self-Efficacy → Achievement ^a	1798.19 (1)	0.90 (1)	.341	.023	.990

^aFit is compared with the Factorial Invariance Across Gender Model

The Configural Invariance Model showed excellent fit to the data suggesting the model is an accurate representation of the observed data χ^2 (124) =1766.09, p <.001), RMSEA =0.024, and CFI =.991. When observed cross-group construct differences exist, those differences are likely due to differences between the relationships between the constructs, not to fundamental measurement differences (Little, 2013). As with the other invariance tests, while there were

statistically significant changes in χ^2 , the CFI and RMSEA changes were less than .01 with the additional constraints, and this suggests that the assumption of factorial invariance is reasonable (Cheung & Rensvold, 2002; Little, 2013). The large sample size is likely the cause of the statistically significant change in χ^2 (Little, 2013).

Table 8. *Unstandardized and Standardized Regression Weight Estimates by Gender*

Measure	Female Students			Male Students		
	UnStd. (S.E.)	Std.	<i>p</i>	UnStd. (S.E.)	Std.	<i>p</i>
9th Grade self-efficacy → 11th Grade self-efficacy	.229 (.018)	.210	<.001	.261 (.017)	.249	<.001
9th Grade identity → 11th Grade identity	.605 (.018)	.548	<.001	.531 (.017)	.506	<.001
9th Grade achievement → 11th Grade achievement	.779 (.009)	.776	<.001	.791 (.008)	.775	<.001
9th Grade self-efficacy → 11th Grade identity	.039 (.020)	.028	.058	.062 (.019)	.048	.001
9th Grade self-efficacy → 11th Grade achievement	.275 (.181)	.018	.128	.518 (.185)	.031	.005
9th Grade identity → 11th Grade self-efficacy	.167 (.015)	.193	<.001	.154 (.014)	.181	<.001
9th Grade identity → 11th Grade achievement	.823 (.150)	.067	<.001	1.236 (.158)	.091	<.001
9th Grade achievement → 11th Grade self-efficacy	.006 (.001)	.081	<.001	.009 (.001)	.144	<.001
9th Grade achievement → 11th Grade identity	.013 (.001)	.148	<.001	.016 (.001)	.197	<.001

Note. UnStd.=Unstandardized regression coefficient; Std.=Standardized regression coefficient, S.E.=Standard Error

Structural Model

The second hypothesis- that although female students are predicted to show lower mathematical identity and self-efficacy than male students, the temporal relation among mathematical self-efficacy and mathematics achievement will not differ between male students and female students were tested (Figures 9 for female students and figure 10 for male students). Although female students did show lower self-efficacy and identity on all items making up these constructs, this was also partially supported. A follow up test of moderation was conducted for all paths by constraining corresponding unstandardized paths across the genders to be the same. Model fit degraded according to a significant Δ , and thus gender showed a moderating impact on the relationship between 9th grade mathematics achievement and 11th grade mathematical self-

efficacy, 9th grade mathematics achievement and 11th grade mathematical identity, and 9th grade and 11th grade mathematical identity (see Table 7). As predicted, the moderators did suggest that self-beliefs will be affected more strongly by achievement in male than in female students. All other relationships did not significantly differ due to gender. Unstandardized and standardized regression weight estimates are reported in Table 8.

The first significant $\Delta \chi^2$ was found when examining the path of 9th grade mathematics achievement to 11th grade mathematical self-efficacy ($\Delta \chi^2 [1] = 8.733, p = .003$) which can be seen in Table 7. This finding suggests that gender moderates the relationship between 9th grade mathematics achievement and 11th grade mathematical self-efficacy. More specifically, it appears that 9th grade mathematics achievement affects 11th grade mathematical self-efficacy slightly more for male students than for female students (see Table 7).

Figure 9. *Female Student Structural Model*

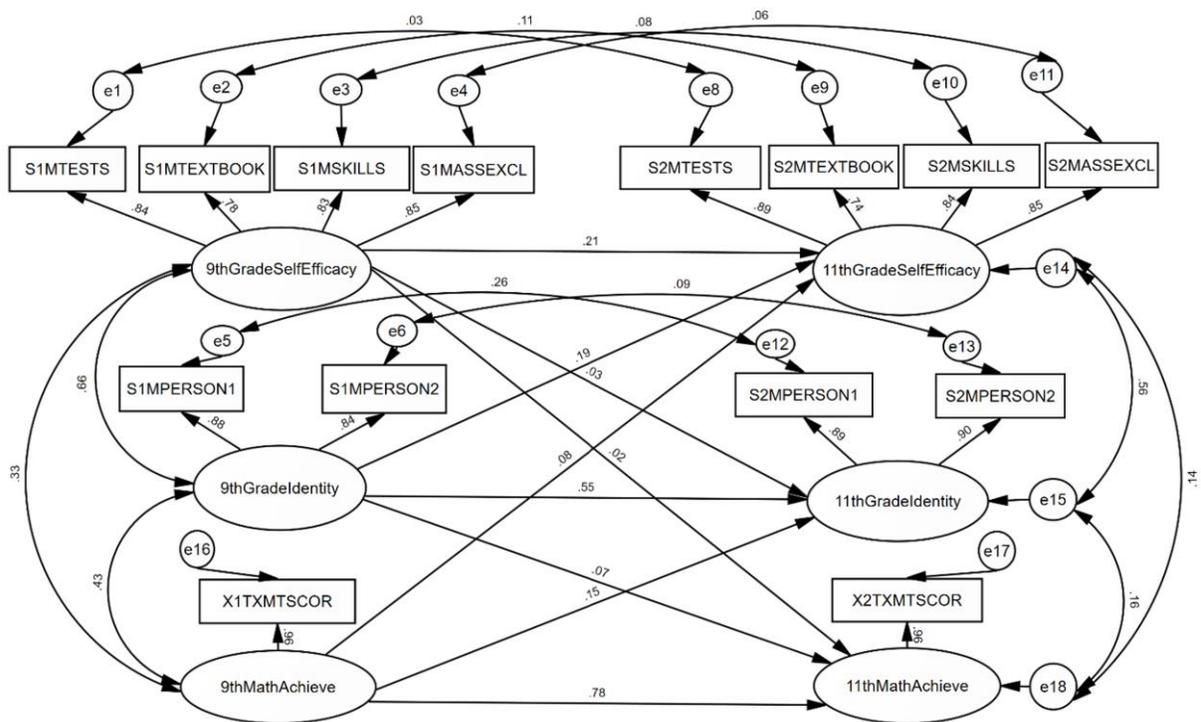
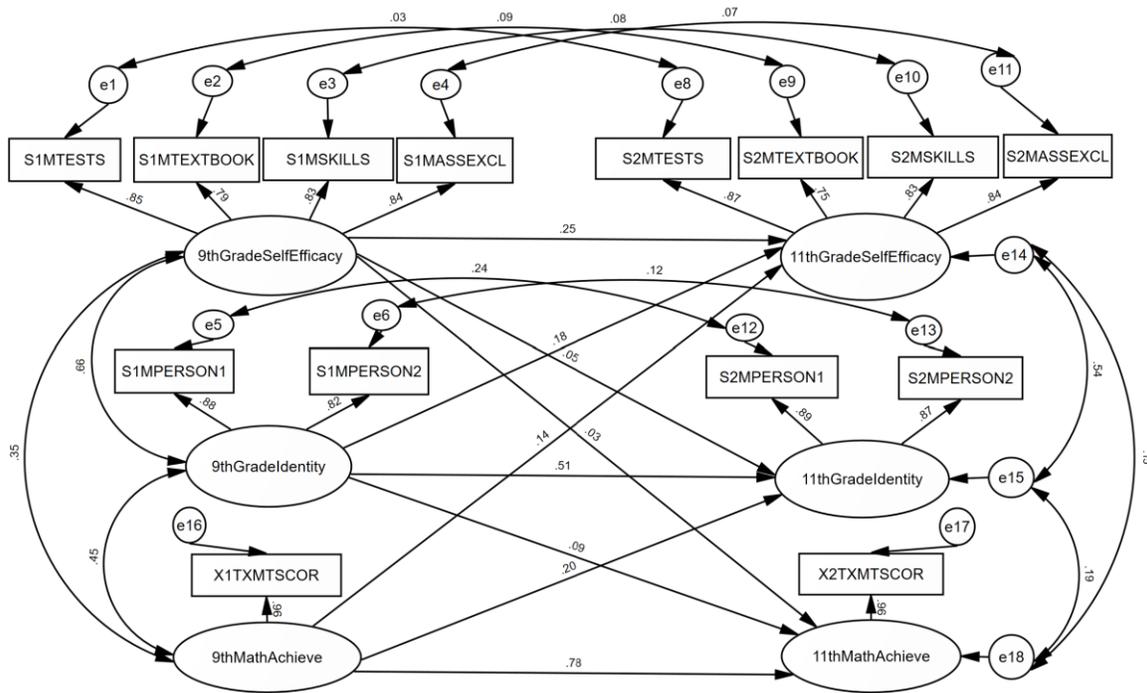


Figure 10. Male Student Structural Model



The next significant $\Delta \chi^2$ was found when examining the path of 9th grade mathematics achievement and 11th grade mathematical identity ($\Delta \chi^2 [1] = 4.39, p = .036$), which can be seen in Table 7. This finding suggests that gender moderates the relationship between 9th grade mathematics achievement and 11th grade mathematical self-efficacy. More specifically, it appears that 9th grade mathematics achievement affects 11th grade mathematical identity slightly more for male students than for female students (see Table 7).

Next, the relationship between 9th grade and 11th grade mathematical identity was tested. To test the $\Delta \chi^2$, the structural paths for 9th grade identity to 11th grade identity were set to equal across male and female student models. The results of the test did show a statistically significant $\Delta \chi^2$ value ($\Delta \chi^2 [1] = 7.30, p = .007$), which can be seen in Table 7. This finding suggests that gender moderates the relationship between 9th grade and 11th grade mathematical identity. More

specifically, mathematical identity is slightly more stable for female students from 9th grade to 11th grade, suggesting that mathematical identity is more sustained and consistent for them.

Finally, it was noted that while there was a significant relationship between 9th grade mathematical self-efficacy and 11th grade mathematical identity, and 9th grade mathematical self-efficacy and 11th grade mathematics achievement in the total sample, these relationships were too small to be meaningfully interpreted. They were likely only significant due to sample size (Gignac & Szodorai, 2016). Also, there was a statistically significant relationship between these factors in male students, but not in female students when the models were run by gender. But, again, the effect sizes were too small to be meaningful. The results of these tests for a moderator did not result in a significant $\Delta \chi^2$ statistic for 9th grade self-efficacy to 11th grade identity ($\Delta \chi^2 [1] = .84, p = .359$) or 9th grade self-efficacy to 11th grade achievement ($\Delta \chi^2 [1] = .91, p = .341$). These results can be seen in Table 7.

Chapter 5: Discussion

The purpose of this study was to investigate the longitudinal relationships between mathematical self-efficacy, mathematical identity, and mathematics achievement, with an emphasis on understanding how these operate in high school students and whether the longitudinal effects work in the same ways for male students as they do for female students. The question of the direction of causality in the relationship between self-beliefs and academic performance has been described as one of the “thorniest issues” in education research (Pajares & Schunk, 2001, p. 239). Two important self-belief constructs that have recently gained the interest of policymakers and educators are mathematical identity and mathematical self-efficacy (Trujillo & Tanner, 2014). This chapter summarizes the results and offers an interpretation of the main findings of this study. The results of the analyses are explored in terms of the context of the broader research literature on mathematics achievement, mathematical self-efficacy, and mathematical identity. The implications for practice, limitations of the current study, and future research directions are discussed.

Summary of Findings

Preliminary Analyses

The purpose of the preliminary analyses was designed to create the basic multivariate cross-lagged panel model that was used to analyze Questions 1 and 2. The central goal of these questions was to understand better the temporal relationship among the key variables, mathematical identity, mathematical self-efficacy, and mathematics achievement. Tests of invariance were conducted and suggested that the constructs of mathematical self-efficacy, mathematical identity, and mathematics self-efficacy all were the same across time and gender, allowing for the conclusion that any changes observed resulted from a change in the measured

factor and not the result of measurement error within the items that make up the factor (Little, 2013).

Hypothesis 1

Based on the literature review, the first hypothesis was that there would be a significant temporal relationship among each of the key variables demonstrated by reciprocal longitudinal effects between mathematical self-efficacy, mathematical identity, and mathematics achievement. The nature of the temporal relationship was expected to be positive between mathematical self-efficacy and mathematics achievement, between mathematical self-efficacy and mathematical identity, and between mathematical identity and mathematics achievement. It was also thought that the strength of the relationship between mathematical identity and mathematical self-efficacy might result in a weaker relationship between mathematical identity and mathematics achievement. In contrast to the findings of the literature review, only parts of this hypothesis were supported.

While the literature review suggested that there would be a significant temporal relationship among each of the key variables, this result was only partially supported. Although there was a significant relationship found between all of the variables, in two instances, the statistical significance appeared to have been the result of the very large sample size and not an accurate indication of a meaningful relationship. The relationships between 9th grade self-efficacy and 11th grade mathematical identity, and also 9th grade self-efficacy and 11th grade achievement both showed path coefficients too small to be interpreted.

Also, the hypothesis that the strength of the relationship between mathematics achievement and mathematical self-efficacy may result in a weaker relationship between mathematical identity and mathematics achievement was supported. The relationship between 9th

grade mathematical identity and mathematics achievement in the measurement model was $r = .43$, while the relationship in the structural model was $\beta = .08$. Therefore, showing that when self-efficacy was included in the model, the strength of the relationship between mathematics achievement and mathematical identity was reduced.

Despite this, self-efficacy was found to be relatively unstable in this population ($\beta = .236$). In comparison, mathematics achievement showed high stability from 9th to 11th grade ($\beta = .776, b = .786, p < .001$) and mathematical identity showed moderately high stability ($\beta = .526, b = .566, p < .001$). This lower correlation suggested that mathematical self-efficacy appears to be somewhat less salient and stable for students from 9th to 11th grade, even after controlling for mathematics achievement and mathematical identity. This variable may be more unstable because individual students may be changing their self-efficacy beliefs at higher rates from 9th to 11th grade, as high school is a key age for the development of self-beliefs (Klimstra, 2010). However, it is also important to note that while the variable showed instability, it did not, on average, show a large mean change. Therefore, while there was some instability within the population, overall, average levels of self-efficacy did not change substantially from 9th to 11th grade.

Hypothesis 2

While the possibility of gender serving as a moderator presented engaging research opportunities, these findings were not found in any of the studies in the literature review, which is why moderation was not expected between mathematical self-efficacy and mathematics achievement. Despite the findings in the literature review, some moderation was detected. Gender showed a moderating effect on the relationship between 9th grade mathematics achievement and 11th grade mathematical self-efficacy, between 9th grade mathematics

achievement and 11th grade mathematical identity, and between 9th grade and 11th grade mathematical identity. In the first two instances, these findings suggested that the 9th grade variable influenced male students' 11th grade variable more than in female students. For example, 9th grade mathematics achievement influenced 11th grade mathematical self-efficacy more for male students than for female students. However, in the third moderation finding, identity was found to be more stable from 9th to 11th grade for male students than for female students.

It is important to note that while tests of χ^2 did show statistically significant results and suggested that gender is a moderator, some of the moderation findings may differ from previous studies due to the statistical power attributed to the large sample in this study. While the differences were statistically significant, the moderating effect was relatively small.

Implications for Theory

Mathematical Self-Efficacy and Mathematics Achievement

This study had several exciting theoretical findings. First, the results support the skill-development model of self-efficacy, as there was no meaningful relationship found between 9th grade mathematical self-efficacy and 11th grade mathematics achievement; however, there was a relationship found between 9th grade mathematics achievement and 11th grade mathematical self-efficacy. The relationship between 9th grade mathematics achievement and 11th grade mathematical self-efficacy showed a moderate effect size. This suggests that while experiencing achievement success in mathematics can develop positive self-belief, having a positive self-belief does not necessarily result in stronger achievement. These findings are consistent with the studies done by Ganley and Lubienski (2016) and Skaalvik and Valås, (1999).

Unlike past longitudinal studies which did not show gender as a moderator in the relationship between mathematical self-efficacy and mathematics achievement (Ganley & Lubienski, 2016), a moderator relationship was found between 9th grade mathematics achievement and 11th grade mathematical self-efficacy. Therefore, these findings show that 9th grade mathematics achievement affects 11th grade mathematical self-efficacy more for male students than for female students. This suggests possible differences in the way that the skill development model affects female and male students. While these differences are small and were likely highlighted due to the large sample size and power of the present study, the Ganley and Lubienski (2016) study was also done using a large sample size, using the Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999. This suggests the possibility that these differences may not develop until students are in high school.

These findings may also be contextualized within the findings of Yee & Eccles (1988) and Marsh (1990). It could be theorized that female students may be more influenced by social comparisons in their development of mathematical self-efficacy and identity. Also, it appears that female students tend to receive feedback that tells them that their accomplishments are due to hard work, rather than due to inherent abilities. These factors combined may result in female students not hearing the social feedback that they need to develop a strong sense of mathematical self-efficacy and mathematical identity, despite receiving equivalent grades to male students.

Mathematical Identity and Mathematics Achievement

Next, the findings of this study can inform researchers about the relationship between mathematical identity and mathematics achievement. Other than one study that found evidence of a small positive relationship between the two (Williams, Burt & Hilton, 2016), the empirical relationship between mathematical identity and mathematics achievement is largely unknown.

There were no studies found in the literature review that explored the temporal relationship between mathematical identity and mathematics achievement. Unlike with self-efficacy, the results found in this study suggest a reciprocal relationship between mathematical identity and mathematics achievement. In other words, it was found that 9th grade mathematical identity can influence 11th grade mathematics achievement, and also 9th grade mathematics achievement can influence 11th grade mathematical identity. However, like self-efficacy, it appears that 9th grade mathematics achievement is a stronger predictor of 11th grade mathematical identity than 9th grade mathematical identity is a predictor of 11th grade mathematics achievement. The relationship between 9th grade achievement and 11th grade identity was moderately strong, whereas the relationship between 9th grade identity and 11th grade achievement was relatively weak.

When a student has high achievement in mathematics, they are more likely to identify with being a math person. High mathematics achievement may increase interest and also influence the person's self-perception. Likewise, when a person identifies as a math person, they are more likely to have high achievement in mathematics, as they are more likely to engage in mathematics activities, consistent with their identity (Grootenboer, Zevenbergen & Chinnappan, 2008). This study's findings suggest that a 9th grader experiencing success in mathematics achievement is more likely to result in the 11th grader identifying with math. Also, a 9th grader identifying with math does appear to explain them achieving in mathematics in 11th grade, though the explanation is weak. Thus, this study confirms a theoretical reciprocal and mutually beneficial relationship between the two over time.

Gender was also found to moderate the relationship between 9th grade mathematics achievement and 11th grade mathematical identity. More specifically, 9th grade mathematics

achievement appeared to more strongly influence 11th grade mathematical identity for male students than for female students. This finding suggests differences in the way that experiencing success in mathematics affects male and female student mathematical identity development.

Finally, 9th grade mathematical identity development had a more substantial influence on 11th grade identity development for female students, suggesting that mathematical identity appears to be more stable for these students. Given these findings, it is possible that female students develop the level of mathematical identity that they are going to experience already by 9th grade. Therefore, their experiences with mathematics achievement do not as significantly influence and result in a change in their mathematical identity by 11th grade. In contrast, male students experience more instability in their mathematical identity scores, which also seem to be influenced more by their experience of higher achievement in mathematics.

Mathematical Identity and Mathematical Self-Efficacy

This study provides additional information about the relationship between mathematical identity and mathematical self-efficacy. While there were no longitudinal studies found in the literature review, some studies did examine the correlations between the variables, and the variables showed a strong correlation, even controlling for utility and interest in mathematics (Middleton, 2013). These results were replicated with this study, which found that at each time point, mathematical self-efficacy and mathematical identity were strongly related, but also distinct concepts.

Interestingly, the results of the present study suggest that 9th grade mathematical identity influences 11th grade mathematical self-efficacy, but 9th grade mathematical self-efficacy does not influence 11th grade mathematical identity. A 9th grader who more strongly identifies as a math person is more likely to have higher self-efficacy in 11th grade, even after controlling for 9th

grade mathematical efficacy, mathematical identity, and mathematics achievement. However, having stronger belief in mathematical abilities in 9th grade is not likely to result in increases in the likelihood of an 11th grader identifying as a mathematician. Therefore, someone who identifies as a mathematician is more likely to believe they are competent in mathematics later on, but someone who experiences belief in their competency in mathematics is not more likely to identify as a mathematician later on.

Limitations

The first limitation is inherent within studies that use data gathered by someone other than the researcher examining the data. While the National Center for Education Statistics is a highly respected distributor of education data, there are still always questions about how the data were gathered and if the data is truly representative of the students that were studied. Also, studies using a large sample size can be subject to a lot of power. This is one possible explanation for the finding of gender as a moderator, despite not observing these findings in the literature review. Studies looking to replicate these findings may need an equally large sample size to find gender to be a moderator.

The fact that the data were not explicitly gathered for this study leads to the second limitation. While the items showed good reliability, strong factor loadings, and good fit with the mathematics achievement, mathematical self-efficacy, and mathematical identity constructs, these items may not have been the best estimate of the constructs. There may be elements of the mathematical self-efficacy and mathematical identity constructs that were not measured in the assessment. Also, this dataset only provides data on the key variables for students in 9th and 11th grade, offering only two time points.

Another limitation is that society's understanding of sex and gender has evolved since 2009 when the dataset was initially gathered. The students' gender data used in this study was a composite variable taken from the base year student questionnaire, parent questionnaire, and school-provided sampling roster. On each of these assessments, the student or parent was asked to indicate whether the student was male or female. At the base year, if the gender indicated by any of the three sources was inconsistent, a researcher manually reviewed the sample member's first name and selected the sex they thought best matched the name. This was corrected at the follow up, which is the dataset being used here. At that time, the variable was updated when missing with data obtained from the student's questionnaire, on which they were asked to indicate male or female (Ingels et al., 2011; Ingels et al., 2013).

Given how the sex/gender variable was assessed, it would be impossible to know if the people completing the test were referring to the student's biological sex or gender identity. However, this distinction is likely not useful anyway. Throughout this dissertation, sex/gender have been used interchangeably. Since the 1970s, psychologists have attempted to highlight distinctions between "sex" referring to biological differences, and gender referring to cultural distinctions and how the person identifies. This division has become less meaningful and is currently dismissed by many modern psychologists, as it has become more evident to sex/gender scholars that biological and cultural forces are inevitably combined and influence human behavior in ways that cannot truly be divided. A clean break between the concept of biological sex and cultural gender identity assumes that we can neatly separate the biological makeup of a person from the impact of existing within society and culture. This hypothesis of the value of dividing these variables can be questioned particularly, concerning transgender, intersex, and nonbinary identities (Donaghue, 2018; Hyde, Bigler, Joel, Tate, & Van Anders, 2019). While

distinguishing between sex/gender is likely not a limitation of the dataset, there is a clear limitation in the dataset limiting out people who identify as neither male nor female. In forcing the participants to choose, they have eliminated the potential for studying and understanding students that do not identify with one of those categories.

Implications for Future Research and Practice

Future researchers can address some of the limitations of this study by examining different datasets that target different ages to understand the relationship between the key variables in students of other ages. They may also examine a dataset that offers three or more time points, as this will provide additional insight into the temporal relationship between the variables. Future researchers should also seek to replicate these findings with regard to mathematical identity, as there is little previous longitudinal research examining this construct. This study may provide particularly relevant future direction to researchers interested in mathematical self-efficacy, which is a popular construct to study. Those researchers may seek to replicate the finding that 9th grade mathematical identity can influence 11th grade mathematical self-efficacy.

Also, researchers may want to examine data sets that recognize gender and sex as a spectrum. This will allow a better understanding of people who are intersex, nonbinary, transgender, or otherwise do not identify as male or female, which can help to develop a more nuanced understanding of these variables and their relationship to sex, gender, and gender identity. They may also want to examine mathematical self-efficacy, mathematical identity, and mathematics achievement using datasets that offer additional items or methods to define these variables. This could help to understand these constructs better. Also, it is essential to recognize the intersections that gender may have with race and ethnicity. Adding in these variables would

help to elucidate further the potential nuanced relationships between race, ethnicity, gender, and the development of self-beliefs and achievement in mathematics.

These findings can guide interventions to promote the key variables of interest, mathematical identity, mathematical self-efficacy, and mathematics achievement. The results suggest that interventions in mathematical identity and mathematical self-efficacy will not substantially increase male or female student mathematics achievement when controlling for earlier levels of the constructs. Instead, educators and interventionists would be best served to intervene in ways that promote mathematics achievement, recognizing that increases in mathematics achievement may also increase mathematical self-efficacy and identity. Also, the best predictor of 11th grade mathematics achievement, mathematical self-efficacy, and mathematical identity, was the levels of these variables in 9th grade. Therefore, earlier intensive interventions that help to develop mathematics achievement before 9th grade would likely promote these variables in high school age students.

Despite this, these findings also tentatively suggest that different interventions may work differently for male and female students, as 9th grade mathematics achievement influenced both 11th grade mathematical identity and mathematical self-efficacy more for male students than for female students. However, identity was more stable for female students from 9th to 11th grade. That being said, the effect size differences between male and female students were small and may have been only statistically significant due to the large sample size. Therefore, limited conclusions might be made for policy and educational implications from these differences. More research is needed to draw practical conclusions from these small results.

Conclusions

This study contributed several key findings. First, this study contributed to the development of theories around mathematical identity, which is a construct that has not been widely studied using quantitative methods. This study suggests that there is a reciprocal relationship between mathematical identity and mathematics achievement in 9th to 11th grade students. The study also suggests a one-way relationship between mathematical identity and mathematical self-efficacy, with mathematical identity influencing mathematical self-efficacy. Next, this study supported the skill development model of self-efficacy, showing that from 9th to 11th grade, mathematics achievement influences mathematical self-efficacy, but mathematical self-efficacy does not influence mathematics achievement. Finally, while additional research is needed, this study will allow policymakers and educators to begin to understand better the path male and female students follow to succeed in mathematics. Specifically, 9th grade female students tend to develop a more stable mathematical identity that is less affected by mathematics achievement success. In contrast, 9th grade male students appear to be more affected by mathematics achievement success in their development of both mathematical identity and mathematical self-efficacy.

References

- Anderman, E. M., Maehr, M. L., Midgley, C. (1999). Declining motivation after the transition to middle school: Schools can make a difference. *Journal of Research and Development in Education, 32*(3), 131–147.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. NY: W.H. Freeman and Company.
- Ball, D. L. (2003). *Mathematical proficiency for all students: Toward a strategic research and development program in mathematics education*. RAND Corporation. Santa Monica, CA. Corporation. https://www.rand.org/pubs/monograph_reports/MR1643.html.
- Baraldi, A.N., & Enders, C.K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48*, 5-37.
- Baumeister, R. (2005). Rethinking self-esteem: why nonprofits should stop pushing self-esteem and start endorsing self-control. *Stanford Social Innovation Review Winter, 2005*, 1-9.
- Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles?. *Psychological Science in the Public Interest, 4*(1), 1-44.
- Blickenstaff, J. C. (2005). Women and science careers: leaky pipeline or gender filter? *Gender and Education, 17*, 369–386.
- Bong, M. (1998). Tests of the internal/external frames of reference model with subject-specific academic self-efficacy and frame-specific academic self-concepts. *Journal of*

- Educational Psychology*, 90(1), 102-110.
- Bong, M., & Clark, R. E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist*, 34(3), 139-153.
- Bong, M., Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15, 1–40.
- Botella, C., Rueda, S., López-Iñesta, E., & Marzal, P. (2019). Gender diversity in STEM disciplines: A multiple factor problem. *Entropy*, 21(1), 30.
- Brainard, S. G. & Carlin, L. (1998). A six-year longitudinal study of undergraduate women in engineering and science, *Journal of Engineering Education*, 87(4), 17–27.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, 9(2), 233-255.
- Donaghue, N. (2018). Discursive psychological approaches to the (un) making of sex/gender. In N. Dess, J. Marecek, L. Bell (Eds.) *Gender, sex, and sexualities: Psychological perspectives*, 127-146. New York, NY: Oxford University Press
- Darragh, L. (2016). Identity research in mathematics education. *Educational Studies in Mathematics*, 93(1), 19-33.
- Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L. (1983). Expectations, values and academic behaviors. In J. T. Spence (Ed.), *Perspectives on achievement and achievement motivation*, 75–146. San Francisco, CA: W. H. Freeman.
- Eccles, J. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist*, 44, 78–89.
- 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, 103–127.

- Enders, C. K. (2010). *Applied missing data analysis*. Guilford press.
- Erikson, E.H. (1968) *Youth, Identity and Crisis*. Norton, New York.
- Esmonde, I. (2009). Ideas and identities: Supporting equity in cooperative mathematics learning. *Review of Educational Research*, 79(2), 1008-1043.
- Ferla, J., Valcke, M., Cai, Y. (2009). Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and Individual Differences*, 19, 499–505.
- Finkel, S. E. (1995). *Causal analysis with panel data*. Thousand Oaks, CA: Sage Publications.
- Froschl, M., & Sprung, B. (2016). Organization spotlight: Furthering girls' mathematical identity: The key to girls' mathematical success. *Childhood Education*, 92(4), 320-323.
- Ganley, C. M., & Lubienski, S. T. (2016). Mathematics self-efficacy, interest, and performance: Examining gender patterns and reciprocal relations. *Learning and Individual Differences*, 47, 182-193.
- Ganley, C. M., & Vasilyeva, M. (2011). Sex differences in the relation between mathematical performance, spatial skills, and attitudes. *Journal of Applied Developmental Psychology*, 32(4), 235-242.
- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549–576.
- Greenberg, D. F. (2008). Causal analysis with nonexperimental panel data. In S. Menard (Ed.), *Handbook of longitudinal research: Design, measurement, and analysis*,

259-278. Burlington, MA: Elsevier/Academic press

- Grootenboer, P., Zevenbergen, R. & Chinnappan, M. (2008) Identity as a lens to understand learning mathematics: Developing a model (Proceedings of the 29th annual conference of Mathematics Education Research Group of Australasia, Vol. 2, pp. 612-615). Canberra, Australia.
- Hazari, Z., Sonnert, G., Sadler, P. M., & Shanahan, M. C. (2010). Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study. *Journal of Research in Science Teaching*, 47(8), 978-1003.
- Hazari, Z., Sadler, P. M., & Sonnert, G. (2013). The science identity of college students: Exploring the intersection of gender, race, and ethnicity. *Journal of College Science Teaching*, 42(5), 82-91.
- Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, 49(5), 505-528.
- Huang, Xiaoxia & Cribbs, Jennifer & Piatek-Jimenez, Katrina. (2016). Understanding undergraduates' math-related perceptions. (Proceedings of the 2016 annual conference of Psychology of Mathematics Education Conference). Tucson, US.
- Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist*, 74(2), 171–193.
- Ingels, S.J., Herget, D., Pratt, D.J., Dever, J., Copello, E., and Leinwand, S. (2010). High School Longitudinal Study of 2009 (HSLs:09) Base-Year Field Test Report (NCES 2011-01). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

- Ingels, S.J., Pratt, D.J., Herget, D.R., Burns, L.J., Dever, J.A., Ottem, R., Rogers, J.E., Jin, Y., and Leinwand, S. (2011). High School Longitudinal Study of 2009 (HSL:09). Base-Year Data File Documentation (NCES 2011-328). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Ingels, S.J., Pratt, D.J., Herget, D.R., Dever, J.A., Fritch, L.B., Ottem, R., Rogers, J.E., Kitmitto, S., and Leinwand, S. (2013). High School Longitudinal Study of 2009 (HSL:09) Base Year to First Follow-Up Data File Documentation. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- Keith, T.Z. (2006). *Multiple regression and beyond*. Boston, MA: Pearson Education, Inc.
- Kenny, D. A. (2015). Measuring model fit. Retrieved from <http://davidakenny.net/cm/fit.htm>
- Klimstra, T. A., Hale III, W. W., Raaijmakers, Q. A., Branje, S. J., & Meeus, W. H. (2010). Identity formation in adolescence: Change or stability? *Journal of Youth and Adolescence*, 39(2), 150-162.
- Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling* (Second ed.). New York: Guilford Press.
- Knuth, E. J., Alibali, M. W., McNeil, N. M., Weinberg, A., & Stephens, A. C. (2005). Middle school students' understanding of core algebraic concepts: Equivalence & variable. *Zentralblatt für Didaktik der Mathematik/International Journal on Mathematics Education*, 37, 68-76.
- Kost, L. E., Pollock, S. J., & Finkelstein, N. D. (2009). Unpacking gender differences in students' perceived experiences in introductory physics. (Proceedings of the 2009 American Institute of Physics, 1179(1), 177-180).

- Kotok, S. (2017). Unfulfilled potential: High-achieving minority students and the high school achievement gap in math. *High School Journal, 100*(3), 183-202.
- Lent, R. W., Brown, S. D., & Gore Jr, P. A. (1997). Discriminant and predictive validity of academic self-concept, academic self-efficacy, and mathematics-specific self-efficacy. *Journal of Counseling Psychology, 44*(3), 307–315
- LoGerfo, L., Christopher, E. M., & Flanagan, K. D. (2011). High School Longitudinal Study of 2009 (HSLs: 09). A First Look at Fall 2009 Ninth-Graders' Parents, Teachers, School Counselors, and School Administrators. NCES 2011-355. *National Center for Education Statistics*.
- Little, T.D. (2013). *Longitudinal structural equation modeling*. New York, NY: The Guilford Press.
- Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development, 31*, 357–365
- Marsh, H. W. (1990). The causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology, 82*, 646-656.
- Marsh, H. W., Smith, I. D., & Barnes, J. (1984). Multidimensional self-concepts: Relations with inferred self-concepts and academic achievement. *Australian Journal of Psychology, 36*, 367–386.
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development, 76*(2), 397-416.

- Marshman, E. M., Kalender, Z. Y., Nokes-Malach, T., Schunn, C., & Singh, C. (2018). Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: A cause for alarm?. *Physical Review Physics Education Research*, *14*(2), 1-17.
- Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among US students. *Science Education*, *95*, 877-907.
- Marx, D. M., Stapel, D. A., & Muller, D. (2005). We Can Do It: The Interplay of Construal Orientation and Social Comparisons Under Threat. *Journal of Personality and Social Psychology*, *88*(3), 432-446.
- McGee, E. O., & Martin, D. B. (2011). "You would not believe what I have to go through to prove my intellectual value!" Stereotype management among academically successful Black mathematics and engineering students. *American Educational Research Journal*, *48*(6), 1347-1389.
- Mertler, C. A., & Vannatta, R. A. (2005). *Advanced and multivariate statistical methods* (3rd ed.): Glendale, CA.
- Middleton, J. A. (2013). More than motivation: The combined effects of critical motivational variables on middle school mathematics achievement. *Middle Grades Research Journal*, *8*, 77-95.
- Middleton J.A., Mangu D., Lee A. (2019) A Longitudinal Study of Mathematics and Science Motivation Patterns for STEM-Intending High Schoolers in the US. In: Hannula M., Leder G., Morselli F., Vollstedt M., Zhang Q. (Eds.) *Affect and Mathematics Education*. ICME-13 Monographs. Springer, Cham.

- Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research, 79*(3), 1129-1167.
- Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*(1), 30-38.
- National Center for Education Statistics (2011). The Nation's Report Card: Mathematics 2011(NCES 2012-458). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.
- National Research Council. (1998). *The nature and role of algebra in the K-14 curriculum*. Washington, DC: National Academy Press.
- 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*(2), 193-203.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4), 543-578.
- Pajares, F., Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology, 24*, 124-139.
- Pajares, F., & Schunk, D. H. (2001). Self-beliefs and school success: Self-efficacy, self-concept, and school achievement. In R. J. Riding & S. G. Rayner (Eds.), *International perspectives on individual differences, Vol. 2. Self perception*, 239-265. Westport, CT, US: Ablex Publishing.

- Palmer, R.F., & Royall, D.R. (2010). Missing data? Plan on it! *Journal of the American Geriatric Society*, 58, 343-348.
- Pomerantz, E. M., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance. *Journal of Educational Psychology*, 94, 396–404.
- Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived mathematical effort expenditure. *British Journal of Educational Psychology*, 84(1), 152–174
- Radovic, D., Black, L., Williams, J., & Salas, C. E. (2018). Towards conceptual coherence in the research on mathematics learner identity: a systematic review of the literature. *Educational Studies in Mathematics*, 99(1), 21-42.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353–387.
- Robnett, R. D., Chemers, M. M., & Zurbriggen, E. L. (2015). Longitudinal associations among undergraduates' research experience, self-efficacy, and identity. *Journal of Research in Science Teaching*, 52(6), 847-867.
- Roth, P.L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, 47, 537-560.
- Schafer, J.L., & Graham, J.W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147-177.
- Scherrer, V., & Preckel, F. (2019). Development of Motivational Variables and Self-Esteem

- During the School Career: A Meta-Analysis of Longitudinal Studies. *Review of Educational Research*, 89(2), 211-258.
- Schunk, D. H., Hanson, A. R., & Cox, P. D. (1987). Peer-model attributes and children's achievement behaviors. *Journal of Educational Psychology*, 79(1), 54-61.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi Experimental Designs*. New York City, New York: Houghton Mifflin Company.
- Shell, D. F., Colvin, C., Bruning, R. H. (1995). Self-efficacy, attribution, and outcome expectancy mechanisms in reading and writing achievement: Grade-level and achievement-level differences. *Journal of Educational Psychology*, 87, 386–398.
- Skaalvik, E. M., & Rankin, R. J. (1995). A test of the internal/external frame of reference model at different levels of mathematical and verbal self-perception. *American Educational Research Journal*, 35, 161–184.
- Skaalvik, E. M., & Valås, H. (1999). Relations among achievement, self-concept, and motivation in mathematics and language arts: A longitudinal study. *The Journal of Experimental Education*, 67(2), 135-149.
- Spinath, B., Spinath, F. M. (2005). Longitudinal analysis of the link between learning motivation and competence beliefs among elementary school children. *Learning and Instruction*, 15, 87–102.
- Stinson, D., & Bullock, E. (2012). Critical postmodern theory in mathematics education research: A praxis of uncertainty. *Educational Studies in Mathematics*, 80, 41–55.
- Talsma, K., Schüz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences*, 61, 136-150.

- Trujillo, G., & Tanner, K. D. (2014). Considering the role of affect in learning: Monitoring students' self-efficacy, sense of belonging, and science identity. *CBE—Life Sciences Education, 13*(1), 6-15.
- Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Self-esteem, academic self-concept, and achievement: how the learning environment moderates the dynamics of self-concept. *Journal of Personality and Social Psychology, 90*(2), 334-349.
- Tyagi, T. K., & Singh, B. (2014). The application of cross-lagged panel analysis in educational research. *Facta Universitatis, Series: Philosophy, Sociology, Psychology and History, 39*-51.
- Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relations between self-beliefs and academic achievement: A systematic review. *Educational Psychologist, 39*, 111-133.
- Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J., Freedman-Doan, C., Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology, 89*, 451-469.
- Wigfield, A., & Wagner, A. L. (2005). Competence, motivation, and identity development during adolescence. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation*, 222-239. New York: The Guilford Press.
- Williams, K. L., Burt, B. A., & Hilton, A. A. (2016). Math achievement: a role strain and adaptation approach. *Journal for Multicultural Education*.
- Yee, D. K., & Eccles, J. S. (1988). Parent perceptions and attributions for children's math achievement. *Sex Roles, 19*, 317-333.
- Young, R., & Johnson, D.R. (2015). Handling missing values in longitudinal panel data with

multiple imputation. *Journal of Marriage and Family*, 77, 277-294.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82-91.