Gender Differences in Math Performance Across Development: Exploring the Roles of Anxiety, Working Memory, and Stereotype Threat

Author: Colleen Marie Ganley

Persistent link: http://hdl.handle.net/2345/1982

This work is posted on eScholarship@BC, Boston College University Libraries.

Boston College Electronic Thesis or Dissertation, 2011

Copyright is held by the author, with all rights reserved, unless otherwise noted.
Gender Differences in Math Performance Across Development: Exploring the Roles of Anxiety, Working Memory, and Stereotype Threat

Dissertation

by

Colleen M. Ganley

submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

August 2011
Abstract

Gender Differences in Math Performance Across Development:

Exploring the Roles of Anxiety, Working Memory, and Stereotype Threat

Colleen M. Ganley

Dissertation Chair: Marina Vasilyeva

This research explored the nature of gender differences in math performance across development. It examined potential mechanisms underlying gender differences by testing a mediation model in which females’ higher anxiety taxes their working memory resources leading to underperformance on a mathematics test. Further, this research examined stereotype threat effects on math performance by testing whether female students presented with a scenario activating the stereotype would perform worse than females not exposed to the stereotype.

Participants in Study 1 were 71 fourth, 107 eighth, and 147 twelfth grade students from high performing school districts. Students completed anxiety measures and a challenging mathematics test either in the stereotype threat condition or the no-threat condition. Results showed that there were consistent gender differences in math performance across all three grade levels; however, stereotype threat did not impact girls’ math performance. Importantly, the relation between gender and math performance at the eighth and twelfth grade levels was mediated by the worry component of anxiety. This
finding suggests that girls’ heightened worry can explain their underperformance on a math test.

In Study 2, the mediating relation observed in Study 1 was further explored by testing whether working memory mediated the relation between worry and math performance. Participants were 90 college students who were assigned to either the stereotype threat or no-threat condition. Students completed anxiety measures, two working memory tasks (verbal and visual), and a challenging math test. Again, findings showed a significant gender difference in math performance but no stereotype threat effects. Further, there was a mediating chain from gender to the worry component of anxiety to visual working memory to math performance. The results suggest that females’ heightened worry taxes their working memory leading to gender differences in math performance.

Both studies contribute to our understanding of affective and cognitive factors underlying gender differences in math performance. The findings of this research are discussed in terms of their implications for interventions and the future of women’s participation in STEM careers.
Acknowledgements

I want to gratefully acknowledge the members of my dissertation committee – Dr. Marina Vasilyeva (Chair), Dr. Beth Casey, Dr. Elida Laski, and Dr. Lillie Albert – for their invaluable feedback during the many stages of the dissertation process. I especially would like to thank Marina and Beth for their guidance throughout my years at BC – you have been great models as researchers and mentors and I truly appreciate all that you have done for me.

Thank you to the Boston College Graduate Student Association for supporting this work through an Individual Research Grant. I was also supported while writing my dissertation by the Boston College Lynch School of Education Dissertation Fellowship.

I would like to thank Alana Dulaney and Katie-Beth Price for assistance with data collection and with many other aspects of the project. Thank you both for your help – you made my life so much easier.

I am also grateful to the professors, principals, and teachers who welcomed me into their schools and classrooms as well as the students who participated in this study.

I would never have made it to this point without the loving support of my family – my parents, Marge and Tom Ganley, and my sister, Jennifer Olson. Mom and dad, thank you for always being there for me, throughout these many years of school. Thanks for letting me know I could be anything I wanted to be, and dad, for encouraging me to be an engineer, even though I didn’t listen. Jen, thanks for being everything a sister should be and for providing me with nephews to do psychological experiments with. Thank you all for being model parents and teaching me about parenting and child development in the
“real world.” Thank you (and Gary, Justin, and Liam) for listening to me talk about my dissertation constantly and for traveling to Boston to support me in the final stages of this endeavor. I love you.
# Table of Contents

LIST OF TABLES.................................................................................................................. viii

LIST OF FIGURES................................................................................................................ ix

CHAPTER 1: INTRODUCTION .............................................................................................. 1

Gender Differences in Math Performance: Developmental Trajectory and Underlying Mechanisms .................................................................................................................. 1

Stereotype Threat Effects on Math Performance: Developmental Trajectory and Underlying Mechanisms ................................................................................................. 2

Present Work: Addressing Gaps in Current Understanding of Gender Differences in Math Performance .................................................................................................... 4

CHAPTER 2: LITERATURE REVIEW ..................................................................................... 6

Gender Differences in Math Performance in the United States ........................................ 6

Potential Explanations for Gender Differences in Math Performance ................................ 9

The Joint Role of Anxiety and Working Memory in Gender Differences in Math Performance .................................................................................................................. 11

Defining Key Terms ........................................................................................................... 12

Gender Differences in Anxiety .......................................................................................... 14

Anxiety and Working Memory ......................................................................................... 15

Anxiety, Working Memory, and Math Performance ......................................................... 17

Anxiety and Working Memory as Mediators of Gender Differences in Math Performance ........................................................................................................... 18
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Role of Stereotype Threat in Gender Differences in Math Performance</td>
<td>20</td>
</tr>
<tr>
<td>Stereotype Threat Effects in Childhood and Adolescence: Developmental</td>
<td>23</td>
</tr>
<tr>
<td>Requirements</td>
<td></td>
</tr>
<tr>
<td>Stereotype Threat Effects in Childhood and Adolescence: Empirical</td>
<td>25</td>
</tr>
<tr>
<td>Findings</td>
<td></td>
</tr>
<tr>
<td>Exploring Mechanisms Underlying Stereotype Threat Effects</td>
<td>31</td>
</tr>
<tr>
<td>Stereotype Threat and Anxiety</td>
<td>33</td>
</tr>
<tr>
<td>Stereotype Threat, Anxiety, Working Memory and Math</td>
<td>35</td>
</tr>
<tr>
<td>Performance</td>
<td></td>
</tr>
<tr>
<td>Present Research</td>
<td>37</td>
</tr>
<tr>
<td>Study 1</td>
<td>37</td>
</tr>
<tr>
<td>Study 2</td>
<td>39</td>
</tr>
<tr>
<td>CHAPTER 3: STUDY 1</td>
<td>40</td>
</tr>
<tr>
<td>Method</td>
<td>43</td>
</tr>
<tr>
<td>Participants</td>
<td>43</td>
</tr>
<tr>
<td>Materials</td>
<td>46</td>
</tr>
<tr>
<td>Stereotype Threat Manipulation</td>
<td>46</td>
</tr>
<tr>
<td>State Anxiety</td>
<td>48</td>
</tr>
<tr>
<td>Math Performance</td>
<td>49</td>
</tr>
<tr>
<td>Procedure</td>
<td>51</td>
</tr>
<tr>
<td>Results</td>
<td>52</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>---------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Preliminary Analyses</td>
<td>52</td>
</tr>
<tr>
<td>Normality Check</td>
<td>52</td>
</tr>
<tr>
<td>Outliers</td>
<td>53</td>
</tr>
<tr>
<td>Reliability</td>
<td>53</td>
</tr>
<tr>
<td>Descriptive Statistics</td>
<td>55</td>
</tr>
<tr>
<td>Means and Standard Deviations</td>
<td>55</td>
</tr>
<tr>
<td>Correlations</td>
<td>57</td>
</tr>
<tr>
<td>Main Statistical Analyses</td>
<td>58</td>
</tr>
<tr>
<td>Gender Differences and Stereotype Threat Effects Across Development</td>
<td>59</td>
</tr>
<tr>
<td>Anxiety as a Mediator of Gender Differences in Math Performance</td>
<td>62</td>
</tr>
<tr>
<td>Analytic Strategy</td>
<td>62</td>
</tr>
<tr>
<td>Estimation of the Mediation Model</td>
<td>64</td>
</tr>
<tr>
<td>Anxiety as a Mediator of Stereotype Threat Effects on Math Performance</td>
<td>67</td>
</tr>
<tr>
<td>Discussion</td>
<td>68</td>
</tr>
<tr>
<td>CHAPTER 4: STUDY 2</td>
<td>73</td>
</tr>
<tr>
<td>Method</td>
<td>76</td>
</tr>
<tr>
<td>Participants</td>
<td>76</td>
</tr>
<tr>
<td>Materials</td>
<td>76</td>
</tr>
<tr>
<td>Stereotype Threat Manipulation</td>
<td>77</td>
</tr>
</tbody>
</table>
CHAPTER 5: GENERAL DISCUSSION

Gender Differences in Math Performance

Lack of Stereotype Threat Effects

Worry as a Mediator of the Relation between Gender and Math Performance

Worry and Visual Working Memory as Mediators of the Relation between Gender and Math Performance

Limitations and Future Directions

Implications

REFERENCES

APPENDICES

Appendix A: Fourth Grade Stereotype Threat Manipulation Word Problems

Appendix B: Eighth Grade Stereotype Threat Manipulation Word Problems

Appendix C: Twelfth Grade Stereotype Threat Manipulation Word Problems

Appendix D: Worry Scale

Appendix E: Emotion List

Appendix F: Fourth Grade Mathematics Test Items

Appendix G: Eighth Grade Mathematics Test Items

Appendix H: Twelfth Grade and College Mathematics Test Items
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Summary of Literature on Stereotype Threat in Children and Adolescents</td>
<td>28</td>
</tr>
<tr>
<td>Table 2</td>
<td>Demographic Information, by School</td>
<td>44</td>
</tr>
<tr>
<td>Table 3</td>
<td>Percentage of Students at Each Achievement Level on State Math Test, by School</td>
<td>45</td>
</tr>
<tr>
<td>Table 4</td>
<td>Reliabilities for Measures by Grade</td>
<td>55</td>
</tr>
<tr>
<td>Table 5</td>
<td>Descriptive Statistics (Means and Standard Deviations) by Grade, Gender, and Stereotype Threat Condition</td>
<td>56</td>
</tr>
<tr>
<td>Table 6</td>
<td>Fourth Grade Correlations among Measures</td>
<td>57</td>
</tr>
<tr>
<td>Table 7</td>
<td>Eighth Grade Correlations among Measures</td>
<td>58</td>
</tr>
<tr>
<td>Table 8</td>
<td>Twelfth Grade Correlations among Measures</td>
<td>58</td>
</tr>
<tr>
<td>Table 9</td>
<td>Descriptive Statistics (Means and Standard Deviations) by Gender and Stereotype Threat Condition</td>
<td>82</td>
</tr>
<tr>
<td>Table 10</td>
<td>Correlations among Measures</td>
<td>83</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Predicted Relation between Gender and Math Performance through Anxiety and Working Memory</td>
<td>12</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Predicted Relation between Stereotype Threat and Math Performance through Anxiety and Working Memory</td>
<td>33</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Study 1 Procedure</td>
<td>52</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Math Test Performance Across Gender and Grade</td>
<td>61</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Proposed Mediation Model for the Relation between Gender and Math Test Performance</td>
<td>63</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Results of the Mediation Analysis, Eighth Grade</td>
<td>66</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Results of the Mediation Analysis, Twelfth Grade</td>
<td>67</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Study 2 Procedure</td>
<td>80</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Proposed Mediation Model for the Relation between Gender and Math Test Performance, College Students</td>
<td>86</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Results of the Mediation Analysis, College Students</td>
<td>89</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

Large-scale international studies consistently indicate that U.S. students are underperforming in mathematics and science, considering their educational resources (Fleischman, Hopstock, Pelczar, & Shelley, 2010; Gonzales et al., 2004; Gonzales et al., 2008). Therefore, it has become increasingly important to conduct research on mathematics and the sciences in order to understand the reasons for the underperformance of American students. Female students appear to be at a particular disadvantage in mathematics. In elementary and middle school, gender differences revealing boys’ advantage begin to emerge in some areas of mathematics (e.g., geometry and measurement), and in high school, girls perform worse than boys on a wider range of standardized assessments (College Board, 2009, 2010; Gibbs, 2010; Gonzales et al., 2008; Hyde, Fennema, & Lamon, 1990; Lindberg, Hyde, Petersen, & Linn, 2010; Vasilyeva, Casey, Dearing, & Ganley, 2009). Later on, women are less likely than men to pursue mathematics in college or choose a career in a STEM field, i.e., in science, technology, engineering, or mathematics (National Science Foundation, 2009).

Gender Differences in Math Performance: Developmental Trajectory and Underlying Mechanisms

Despite the large body of literature documenting and exploring gender differences in mathematics, the mechanisms underlying these differences are not completely understood. Researchers have attributed gender differences in mathematics to a number of biological and social factors, including gender socialization (Eccles & Jacobs, 1986). Being socialized as a woman involves exposure to gender stereotypes, including the
stereotype that men are better than women at math. Being aware of these stereotypes may lead girls and women to have heightened anxiety about math testing and to be impacted by the phenomenon of stereotype threat (Hembree, 1988; Steele, 1997).

Indeed, research suggests that girls have higher levels of anxiety about math tests than boys (Hembree, 1988; Hong & Karstensson, 2002). In addition, research shows that anxiety interferes with working memory and that both anxiety and working memory are related to math performance (Eysenck & Calvo, 1992; Hembree, 1988; Raghubar, Barnes, & Hecht, 2010). Thus, given the present state of knowledge, the next logical step involves trying to understand whether gender differences in math test performance can be explained by girls’ heightened anxiety, which leads to working memory deficits and then to poorer math performance. If this is the mechanism at work, knowing this would provide researchers and educators with potential targets for intervention.

**Stereotype Threat Effects on Math Performance: Developmental Trajectory and Underlying Mechanisms**

In the context of discussing affective factors related to gender and math performance, the phenomenon of stereotype threat has been proposed as a potential explanation of the gender effect in math. Stereotype threat refers to a situation in which the negative stereotype about one’s group is activated. According to this theory, people in a stereotype threat situation perform worse on the stereotyped task than those in a situation without threat. The first studies testing this theory focused on racial differences in academic test performance (Steele, 1997; Steele & Aronson, 1995; Steele, Spencer, & Aronson, 2002) but the effect has since been found on women’s math performance,
among other groups (Brown & Josephs, 1999; Schmader, 2002; Spencer, Steele, & Quinn, 1999). For women, there is a commonly recognized stereotype that they are worse than men at math. It has been posited that in a situation where this stereotype is activated, women perform worse on a math test than in a situation involving no threat (Steele, 1997; Spencer, et al., 1999).

There is a considerable amount of literature investigating stereotype threat effects on the math performance of female college students (see Nguyen & Ryan, 2008), but less work has been done on the occurrence of stereotype threat in children and adolescents. The available evidence in this area of research is inconsistent, with some studies finding stereotype threat effects as early as kindergarten and others not finding effects at the high school level (Ambady, Shih, Kim, & Pittinsky, 2001; Huguet & Regner, 2007, 2009; Keller, 2007; Keller & Dauenheimer, 2003; Muzzati & Agnoli, 2007; Neuville & Croizet, 2007; Stricker & Ward, 2004).

When addressing the mechanisms underlying stereotype threat effects, researchers often identify the same cognitive and affective factors that have been associated more generally with gender differences, namely anxiety and working memory. There is limited research evaluating the entire model in which stereotype threat leads to higher anxiety, which leads to working memory deficits, which in turn, results in poorer math performance, but a growing body of work supports many of the bivariate relations involved in this model (e.g., Cadinu, Maass, Rosabianca, & Kiesner, 2005; Crowe, Matthews, & Walkenhorst, 2007; Johns, Inzlicht, & Schmader, 2008; Raghubar, et al., 2010). The paucity of studies examining underlying mechanisms of stereotype threat is
particularly noticeable in developmental research. There is only one study that explores the role of anxiety in stereotype threat effects on mathematics performance during adolescence, and none with working memory or in children (Keller & Dauenheimer, 2003). Thus, although stereotype threat is now a well-established phenomenon, we still do not have a thorough understanding of the development of susceptibility to stereotype threat or the underlying processes involved.

**Present Work: Addressing Gaps in Current Understanding of Gender Differences in Math Performance**

The present dissertation research includes two studies that contribute to the literature by examining the developmental trajectory of gender differences in math performance and the role of anxiety and working memory in this relation. In addition, these studies investigate the development of susceptibility to stereotype threat and explore the role of anxiety and working memory in the impact of stereotype threat on math performance. Given that girls are losing ground compared to boys in mathematics as they get older, it is important to determine at what point girls begin to fall behind and begin to be affected by societal stereotypes. Exploring these issues is not only relevant to developmental theory but also to educational practice. Armed with this information, strategies aimed at reducing stereotype threat can be implemented at the most appropriate age and interventions targeting the affective and cognitive processes underlying gender differences can be conducted.

The main goal of the first study is to investigate affective factors related gender differences in math performance. Gender differences are examined across development
(fourth, eighth, and twelfth grades) and the role of anxiety in this relation is investigated and compared across different ages using parallel measures. Another goal of this study is to investigate the developmental trajectory and underlying mechanisms of stereotype threat effects. Specifically, the study addresses susceptibility to stereotype threat in relation to math performance across school age and the potential role of anxiety in this relation. The second study extends the investigation started in Study 1 by exploring potential mechanisms underlying the relation between anxiety and math performance. This study examines the meditational path through anxiety and working memory as an explanation for both gender differences in math performance and stereotype threat effects on math performance.
Chapter 2: Literature Review

Gender Differences in Math Performance in the United States

There is concern that the United States will be unable to maintain its position as a leader in science and technology due to a shortage of highly-skilled mathematicians and scientists (National Science Board, 2010). This concern is partially fueled by the fact that American students lag behind a number of their international counterparts in mathematics and science despite available educational resources (Fleischman, et al., 2010; Gonzales et al., 2004; Gonzales et al., 2008). Therefore, it has become increasingly important to focus teaching, research, and interventions on mathematics and the sciences for all students. In particular, women are at an additional disadvantage; they perform worse than boys in math as they get older and are underrepresented in math- and science-related careers (NSF, 2009). Until we understand the reasons for this disadvantage and find ways to increase the number of girls in math and science, the future American workforce will not be able to achieve its full potential.

There is currently a debate in the field about whether gender differences in math performance are disappearing. Recent research suggests the gender gap is narrowing; however, there is still evidence of gender differences that vary with age, student ability level, and math complexity (Gibbs, 2010; Hyde, Fennema, & Lamon, 1990; Lindberg, et al., 2010). In their 1990 meta-analysis, Hyde, Fennema, and Lamon found that the total mean effect size for gender was relatively small ($d = 0.20$). Later, in a 2010 meta-analysis, Lindberg and colleagues found that the overall effect size had decreased to a trivial magnitude ($d = 0.07$). However, these general effect sizes mask important gender
differences that emerge when math achievement is examined separately for different
groups of students (based on age and ability level) and for math tasks at different levels of complexity.

Whereas gender differences are negligible during elementary school and into middle school, differences favoring boys appear in high school students ($d = 0.23$) and persist into college ($d = 0.18$) (Lindberg, et al., 2010). In addition, during high school, boys outscore girls on national achievement tests including the quantitative SAT test and the mathematics Advanced Placement (AP) exams (College Board, 2009, 2010). These differences emerge even though girls are taking advanced mathematics courses at a similar rate to boys in high school (College Board, 2010). As girls get older, they tend to drop out of the mathematics pipeline at higher rates. Although girls made up 59% of all college students in 2004, they made up only 45% of college students in mathematics (NSF, 2009). Women are less likely than men to obtain advanced degrees in a mathematical field, being awarded 43% of master’s degrees in mathematics in 2006 and only 30% of doctoral degrees. Additionally, women are less likely to have a career in mathematics and science (NSF, 2009). Since women seem to lag behind men as they get older, it is important to understand the factors related to the progression of the gender gap in mathematics. If girls and boys begin on an even playing field in mathematics, why do girls fall behind?

In addition to the fact that the gender gap in mathematics performance increases as boys and girls get older, there is evidence that gender differences tend to be most pronounced in students of high ability (Halpern et al., 2007; Hedges & Nowell, 1995;
Hyde, Fennema, & Lamon, 1990; Lindberg et al., 2010). In Lindberg and colleagues’ (2010) meta-analysis, the overall effect size for studies of the general population was negligible ($d = 0.07$), but the effect sizes were small to moderate in size in samples of high ability students ($d = 0.15$ for moderately selective samples and $d = 0.40$ for highly selective samples). These results suggest that the magnitude of gender differences increases as the selectivity of the sample increases.

Researchers often find that gender differences exist for only some mathematics content areas and types of problems. Girls perform better than boys in particular areas of math including number and computation skills, which rely heavily on recall of procedures and information (Gibbs, 2010; Hyde, Fennema, & Lamon, 1990; Vasilyeva et al., 2009). However, boys tend to outperform girls in more complex math that involves problem solving, geometry, and measurement (Becker, 1990; Gibbs, 2010; Hyde, Fennema, & Lamon, 1990; Lindberg et al., 2010; Vasilyeva et al., 2009). This fits with the work of Hyde, Lindberg, Linn, Ellis, and Williams (2008), which suggests that gender differences do not exist on state math assessments, which include primarily items of low levels of complexity. These results suggest that different patterns of gender differences emerge on assessments that tap different types of skills.

The three factors that are related to the magnitude of gender differences also interact with one another. First, in relation to the interaction between age and ability level, research suggests that although gender differences do not emerge in high school in the general population, differences emerge as early as middle school for high ability students (Entwisle, Alexander, & Olson, 1994). Second, in regard to the interaction
between age and item complexity, Gibbs (2010) suggests that the increase in the inclusion of complex tasks, on which males excel, in math assessments as students get older can partially explain why gender difference in math increase with age. Gibbs (2010) found that girls outperformed boys on math assessments in preschool due to their advantage in counting and number skills. In addition, girls continued to have an advantage in number skills through the fifth grade. However, in later grades, when more complex skills were required, girls lost ground on the math assessment as boys performed better on these more complex tasks. The male advantage was consistent across some skills (e.g., multiplication and division, $d's \approx 0.10$), but in others, the male advantage increased over time (e.g., rate and measurement, kindergarten $d \approx 0.02$ and 5th grade $d \approx 0.20$). Thus, despite the fact that some researchers emphasize that gender differences in math are small (perhaps disappearing) and emerge only in high school (Lindberg et al., 2010), it appears that if more complex items are included in math assessments, a male advantage may be apparent at a younger age.

**Potential Explanations for Gender Differences in Math Performance**

Researchers have posited a number of potential explanations for gender differences in math performance and for the underrepresentation of women in math- and science-related careers, including both biological factors and social factors (Ceci, Williams, & Barnett, 2009). Investigators focusing on biological explanations have suggested that gender differences in math are due to hormonal influences on the brain (Benbow & Stanley, 1980; Kimura, 1992), genetics (Scarr & Satzman, 1982; Zohar & Guttman, 1988), or evolutionary factors (Geary, 1996). Others believe that social factors,
including gender role socialization (Eccles, 1987; Eccles & Jacobs, 1986) and attributional styles (Heller & Ziegler, 1996) are primarily responsible for gender differences in math. A number of researchers have also proposed that both biological and social factors together explain gender differences in math achievement (Halpern & Tan, 2001; Nuttall, Casey, & Pezaris, 2005).

Social factors, such as gender socialization, may impact math performance through girls’ heightened anxiety about mathematics as well as the situational phenomenon of stereotype threat (Aronson & Steele, 2005; Spencer et al., 1999; Steele, 1997). As girls are socialized as members of their gender, they are inundated with stereotypes about how they should behave. Gender-based stereotypes may lead girls to have higher anxiety about math, due to the fact that they may become familiar with the stereotype that women are not as good as men at math (Eccles & Jacobs, 1986). Exposure to these stereotypes may affect them during math testing situations by leading them to fear confirming the stereotype, which can lead them to underperform (stereotype threat) (Steele, 1997). In support of the socialization theory, researchers find that the degree of gender stereotyping in a society is related to the extent of the gender difference in math in that society. Specifically, Nosek and colleagues (2009) found that nation-level implicit stereotyping about gender and math predicts nation-level gender gaps in math achievement.
The Joint Role of Anxiety and Working Memory in Gender Differences in Math Performance

The integration of a number of bodies of research suggests the possibility that anxiety and working memory play a role in gender differences in math (Hembree, 1988; Osborne, 2001; Raghubar et al., 2010; Williams, 1996). First, there is substantial research demonstrating gender differences in anxiety, in its many manifestations (Hembree, 1988; Miller & Bichsel, 2004; Sowa & LaFleur, 1986; Williams, 1996). Second, there are studies showing that anxiety is related to working memory (Owens, Stevenson, Norgate, & Hadwin, 2008; Rapee, 1993; Shackman et al., 2006). Third, there is evidence that anxiety and working memory are both related to math performance (Deffenbacher, 1980; Hembree, 1988; Holmes & Adams, 2006; Jarvis & Gathercole, 2003; Miller & Bichsel, 2004; Williams, 1996). Combining evidence from these bodies of work raises the possibility that gender differences in math performance may be due to gender differences in anxiety and the relation between anxiety and working memory, as shown in Figure 1. Thus, if girls are more anxious about the math testing situation, their heightened anxiety may interfere with working memory and this interference may lead girls to underperform in math.
Defining key terms. Before reviewing empirical evidence from these bodies of literature in more detail, it is necessary to define some of the main constructs discussed here. Multiple types of anxiety have been studied both in the context of gender differences in anxiety as well as in the context of anxiety’s relation to working memory and math performance (Ashcraft & Kirk, 2001; Hembree, 1988; Miller & Bichsel, 2004; Raghubar, et al., 2010). One major distinction in anxiety research is between the notions of trait and state anxiety. Trait anxiety refers to the general tendency to feel anxious whereas state anxiety refers to temporary feelings of anxiety experienced in a particular situation (Hong & Karstensson, 2002). Research shows that though these different types of anxiety are related to each other (i.e. individuals with high trait anxiety are more likely to have high state anxiety), they represent conceptually separate constructs (Head & Knight, 1988; Hembree, 1990; Miller & Bichsel, 2004).

In addition to distinguishing between state and trait anxiety, it should also be mentioned that anxiety is conceptualized as involving a complex set of reactions...
(cognitive, emotional, behavioral, physiological), which are commonly grouped into two main categories: worry and emotionality (Deffenbacher, 1977; Hembree, 1988; Liebert & Morris, 1967; Sarason, 1984). Worry refers to cognitive reactions including concerns related to performing poorly (Deffenbacher 1977, 1978; Leibert & Morris, 1967). Emotionality refers to the physiological and affective arousal response as well as the awareness of this response (Deffenbacher 1977, 1978; Leibert & Morris, 1967). These dimensions have been validated by factor analyses (e.g., Benson & Tippets, 1990) and have been found to be moderately correlated but to represent separate constructs (Brodish & Devine, 2009; Hembree, 1988; Morris, Davis, & Hutchings, 1981). Research also shows that these components have differential effects on performance (Kim & Rocklin, 1994; Zeidner & Nevo, 1992).

Another construct that is examined in the present study in the context of gender differences in math performance is working memory. Working memory involves temporarily storing information while simultaneously manipulating it (Baddeley & Hitch, 1974). Working memory is a critical skill for academic tasks as well as everyday functioning (Gathercole, Pickering, Knight, & Stegman, 2004). It has been conceptualized as involving three components: the central executive, the visuospatial sketchpad, and the phonological loop. The central executive controls attention and coordinates the functions of the phonological loop (responsible for storage of verbal information) and the visuospatial sketchpad (responsible for the storage of visual/spatial information) (Baddeley & Hitch, 1974). Verbal working memory involves the use of the phonological loop and the central executive whereas visual working memory involves the
visuospatial sketchpad and the central executive (Baddeley & Hitch, 1974). It is likely that both visual and verbal working memory are related to math performance (there has been an ongoing discussion about the extent for each) and thus it is important to investigate both components (e.g., Gathercole, et al., 2004; Holmes, Adams, & Hamilton, 2008).

**Gender differences in anxiety.** Research on anxiety in its various manifestations has consistently found gender differences such that women are more anxious than men both specifically about math as well as more generally about testing situations (Hembree, 1988; Hong & Karstensson, 2002; Hyde, Fennema, Ryan, Frost, & Hopp, 1990; Miller & Bichsel, 2004; Sowa & LaFleur, 1986; Williams, 1996). With regard to the two components of anxiety, most researchers have found that girls are higher than boys in both worry and emotionality (Hembree, 1988; Williams, 1996). One study found that the gender difference was quite a bit larger ($d = 0.64$) for worry compared to emotionality ($d = 0.25$) (Williams, 1996); however, other work has found that gender differences are comparable for the two components (Hembree, 1988).

Research shows gender differences in anxiety across most of development. In Hembree’s meta-analysis (1988) it was found that gender differences in test anxiety were small but significant in the lower elementary years ($d = 0.14$) and increased in the middle elementary years ($d = 0.28$) peaking in grades 5-10 ($d = 0.43$), then declining slightly in high school and college ($d = 0.27$). A similar developmental pattern is found for math anxiety, but differences usually do not develop until the middle school years (Hyde, Fennema, Ryan, et al., 1990).
**Anxiety and working memory.** The relation between anxiety and working memory has been studied extensively in the context of the processing efficiency theory (recently revised as the attentional control theory; Eysenck & Calvo, 1992; Eysenck, Derakshan, Santos, & Calvo, 2007). Findings from this body of literature can prove useful in understanding the nature of the relations that are hypothesized to be involved in gender and stereotype threat effects on math performance. The theory states that individuals with high anxiety perform less efficiently on tasks requiring working memory resources because their worrisome thoughts interfere with their working memory, leading them to be unable to fully utilize their working memory capacity for task performance. This theory is primarily focused on the worry component of anxiety, as it is suggested that it is worry-related thoughts that lead to working memory deficits (Eysenck & Calvo, 1992).

Processing efficiency theory and associated research can help to understand how anxiety and working memory may be related to one another. There are, however, two key distinctions between this research and the research conducted in the present dissertation that must be addressed. First, the processing efficiency theory tends to focus on anxiety as an individual difference variable (trait anxiety). In other words, participants are often grouped based on their general anxiety level and compared with one another to see if the high anxiety group performs worse. Using anxiety in this way may lead to different results when compared to using anxiety as a continuous predictor variable that is more tied to a situation (like state anxiety, which was measured in this study). In addition, this theory focuses more on the effects of anxiety on the amount of time taken to complete
cognitive tasks rather than on the accuracy of performance on these tasks (Eysenck & Calvo, 1992). However, given that in many testing situations, participants are given a limited amount of time to complete the task, slowed processing efficiency may lead to poorer performance because students are unable to complete all of the items (Osborne, 2006). So it is possible that results from some assessments that show poorer female performance really reflect these processing efficiency deficits. Gender differences are often found on timed and untimed tests (Lindberg et al., 2010) and some work has found that stereotype threat impacts both efficiency and accuracy of performance. Thus it is important to consider anxiety’s impact on both of these measures of performance.

Although research generally suggests that anxiety is related to working memory, there are inconsistent results in regard to whether anxiety impacts verbal or visual working memory, or both. Many researchers have found, across development, that anxiety exclusively has its impact on verbal working memory – presumably because intrusive worry-related thoughts may require processing in the phonological loop (Hadwin, Brogan, & Stevenson, 2005; Ikeda, Iwanaga, & Seiwa, 1996; Lee, 1999; Owens et al., 2008; Rapee, 1993). In contrast, other researchers have found that anxiety leads to lower performance on visual working memory tasks, but not on verbal tasks (Crowe, Matthews, & Walkenhorst, 2007; Shackman et al., 2006). However, Miller and Bichsel (2004) found that trait math anxiety was related to visual but not verbal working memory, but state anxiety was not related to either type of working memory. This body of research suggests that anxiety and working memory are likely related to one another, but it is unclear what type of working memory is most affected by anxiety, making it
important to address both types of working memory in research in this area. Potentially the results may be impacted by the different types of anxiety measured as well as the different measures of working memory.

Anxiety, working memory, and math performance. Researchers consistently find that anxiety is related to math performance across a number of anxiety measures and math assessments. This relation has been found with trait anxiety measures (Deffenbacher, 1980; Hembree, 1988; Williams, 1996), math anxiety measures (Ganley & Vasilyeva, under review; Ma, 1999; Miller & Bichsel, 2004) and state anxiety measures in the context of math tests (Beilock, Rydell, & McConnell, 2007; Brodish & Devine, 2009; Cadinu et al., 2005; Osborne, 2001; Spencer et al., 1999). In general, research finds that the relation between anxiety and test performance is stronger for the worry component of anxiety when compared to the emotionality component, however no studies have included the two components in one model (Deffenbacher, 1980; Hembree, 1988; Kim & Rocklin, 1994; Zeidner & Nevo, 1992). Studies comparing the effects of state and trait anxiety have found that the relation between state anxiety and performance is stronger than the corresponding association between trait anxiety and performance (e.g., O’Neil & Fukumura, 1992). Though many studies examining the relation between anxiety and math performance are correlational, there have also been experimental studies designed to alleviate anxiety. These studies have generally found that behavioral and cognitive-behavioral interventions that reduce worry and emotionality also improve performance (in both math and other academic areas), suggesting a causal relation between anxiety and performance (Hembree, 1988; Wood, 2006).
Working memory and math performance are proposed to be related to each other because performing mathematical tasks relies on the ability to hold multiple pieces of information in one’s mind while also processing information, precisely the skills involved in working memory (Raghubar, et al., 2010). Indeed, it has been well established in the literature that working memory is related to mathematics performance. Some studies find evidence that verbal working memory (Adams & Hitch, 1997; Gathercole, et al., 2004; Owens et al., 2008), visual working memory (Holmes, et al., 2008; Kyttala & Lehto, 2008), or both types of working memory are related to performance (Berg, 2008; Holmes & Adams, 2006; Jarvis & Gathercole, 2003; Miller & Bichsel, 2004). This relation has been established in children and adolescents, as well as adults, across a wide range of mathematical tasks. Holmes and Adams (2006) found that the type of working memory related to math performance differed based on students’ age and the characteristics of the mathematics task. In second-grade children, visual working memory, but not verbal working memory, factored into students’ performance on all items. However, for fourth grade children, verbal working memory was related to performance on easy items while visual working memory was related to performance on more difficult items. These results suggest that perhaps the age of students and the variability in mathematics measure used could lead to variability in findings regarding the component of working memory involved in math performance.

Anxiety and working memory as mediators of gender differences in math performance. The combination of findings from these bodies of research raises the possibility that anxiety and working memory serve as mediators of gender differences in
math – that is, gender differences in math may be, in part, due to girls’ higher anxiety, which interferes with working memory, which leads to underperformance. Most researchers have looked only at bivariate relations in this model (e.g., the relation between gender and anxiety). There are a few studies in which investigators looked beyond bivariate relations and started examining mediation (e.g. looking at anxiety as a mediator of the relation between gender and math performance). However, they focus on just one mediator at a time and the full mediation model from gender to anxiety to working memory to performance has never been tested directly.

There are two studies, both of which were conducted with high school students, that analyzed whether anxiety mediated gender differences in math performance (Casey Nuttall, & Pezaris, 1997; Osborne, 2001). The two studies used different conceptualizations of anxiety. Casey and colleagues (1997) examined trait math anxiety (combining emotionality and worry) and found that anxiety did not act as a mediator of the relation between gender and math performance. Osborne (2001) examined state anxiety, specifically the emotionality component measured after testing, and found that it did mediate the gender-math relation. Thus, potentially general anxiety toward math is not as important a factor in gender differences in math when compared to anxiety occurring during the testing situation. It should be noted though that Osborne’s (2001) study, which looked at state anxiety, had a methodological issue because anxiety was measured after performance and several investigators have pointed out that the nature and extent of anxiety may differ following performance (Marx & Stapel, 2006). Thus it is important to examine the relation between anxiety and math performance when
measuring anxiety prior to the test. It is also important to consider both the emotionality and worry components of anxiety as separate variables simultaneously. Further extending this type of research to younger students is key to understanding the developmental trajectory of this relation.

There are no studies in which anxiety and working memory are simultaneously tested as mediators in the context of studying gender differences in math. There is however one recent study (Owens et al., 2008) in which working memory was tested as a mediator of the relation between trait anxiety and math performance. The findings indicated that verbal working memory mediated the relation between trait anxiety and math performance in fifth grade children. However, since gender was not examined as part of their model it is not clear if the same explanatory mechanism can be applied to understanding the nature of gender differences in math performance.

The model tested in the present dissertation expands on this research by investigating whether both anxiety and working memory are part of a mediation model for gender differences in math performance. If this model is operating, it would suggest important targets for potential interventions. Perhaps gender differences in math performance can be decreased if girls’ anxiety is alleviated and their working memory skills are strengthened.

**The Role of Stereotype Threat in Gender Differences in Math Performance**

In discussing the impact of gender socialization on math performance, researchers have often focused on the phenomenon of stereotype threat. Stereotype threat refers to a situation in which a negative stereotype about ones’ group is activated. According to the
theory developed by Steele and colleagues, people in a stereotype threat situation perform worse on the stereotyped task than those in a situation without threat (Steele, 1997; Steele & Aronson, 1995; Steele, et al., 2002). The original paper published about this phenomenon focused on African-American college students and academic test performance (Steele & Aronson, 1995). This pioneering paper found that asking for students’ race prior to an assessment led to underperformance in African-American students. This theory has since been extended to apply to many negatively stereotyped groups, including women on math tests (e.g., Brown & Josephs, 1999; Schmader, 2002; Spencer, et al., 1999), Latino students on academic tests, (e.g., Gonzales, Blanton & Williams, 2002), people of low socioeconomic status on academic tests (e.g., Croizet & Claire, 1998), and the elderly on memory tasks (e.g., Chasteen, Bhattacharyya, Horhota, Tam & Hasher, 2005).

In the case of women in mathematics, there is a commonly recognized stereotype that women are worse than men at math. According to stereotype threat theory, women in a situation where the stereotype is activated perform worse on a math test than women in a situation without threat (Spencer, et al., 1999). Steele (1997) posits that stereotype threat impacts performance because members of the negatively stereotyped group fear confirming the stereotype or that they will be judged based on the stereotype about their group. Steele (1997) suggests that in the long run, stereotype threat effects may lead women to disidentify with and devalue mathematics, which could explain why they are less likely to perform well in mathematics and continue into advanced mathematics. The
important implication is that stereotype threat may lead to women not reaching their full potential in mathematics (Brodish & Devine, 2009).

Steele (1997) and Schmader (2002) have argued that women must be identified with math in order for their math performance to be impacted by stereotype threat. One is considered identified with math if they feel that they are good at math and if they believe that it is important to them to be good at math (Steele, 1997; Smith & White, 2001). Women high in math identification react more negatively to stereotype threat because they need to contend with thoughts suggesting that they should behave in one way (perform poorly) when it is important to them to behave in another way (perform well). Research in high school and college students shows that indeed high math identified women are more susceptible to stereotype threat effects (Good & Aronson, 2008; Keller, 2007; Smith & White, 2001).

In addition to the individual characteristic of math identification, particular aspects of the testing situation are important for stereotype threat effects to occur. In order to induce stereotype threat, the situation needs to be evaluative and gender must be made salient. In order for these two situational factors to lead to stereotype threat effects, the test must also be difficult. Tests that are introduced as ones which are evaluative, or indicative of one’s ability, lead to the feeling that poor performance on the test indicates low ability (Aronson & Steele, 2005; Good & Aronson, 2008; Steele, 1997). This, combined with one’s gender being made salient, leads women to believe that if they perform poorly on the test, they are confirming the stereotype about women being bad at math (Good & Aronson, 2008; Steele, 1997). In addition, stereotype threat effects occur
on difficult tests because women are more likely to perform poorly on these assessments. (Neuville & Croizet, 2007; Nguyen & Ryan, 2008; O’Brien & Crandall, 2003; Spencer et al., 1999; Steele, 1997). Women then think about their poor performance and may feel that their performance confirms the stereotype. It is critical to consider these characteristics of the participants and testing situation when examining stereotype threat.

It is important to note that stereotype threat theory implies that there is some level of stereotype threat occurring during everyday mathematics testing situations (Steele, 1997; Wheeler & Petty, 2001). However, by further activating stereotype threat and comparing this condition to normal testing conditions, we can look at how this intensified threat impacts performance to help us understand the mechanisms involved in the everyday experience of stereotype threat. Thus, though stereotype threat is being manipulated in experimental studies more strongly, this theory suggests that effects occur in normal conditions, but to a lesser extent.

**Stereotype Threat Effects in Childhood and Adolescence: Developmental Requirements**

Stereotype threat is a well-established phenomenon in samples of college women (see Nguyen & Ryan, 2008). However, we do not have a thorough understanding of how susceptibility to stereotype threat develops (Good & Aronson, 2008). Since girls are falling behind boys in math as they get older, it is important to understand at what point they begin to be affected by societal stereotypes. This is important, since stereotype threat potentially contributes to educational and social inequalities, and it is during adolescence (the time that gender differences begin to appear) that identities and future careers
aspirations are developing (Good & Aronson, 2008). Additionally, understanding when stereotype threat occurs can provide evidence to suggest the most appropriate ages at which to target interventions designed to alleviate the effects of stereotype threat.

Aronson and Good (2002) suggest that stereotype threat begins to impact girls’ math performance during early adolescence, when they develop the cognitive and social-cognitive abilities needed to understand the societal and personal implications of negative stereotypes. Aronson and Good (2002) posit four necessary developmental conditions in order for girls’ math performance to be impacted by stereotype threat. Girls must be (1) aware of gender stereotypes, (2) able to understand the societal and personal implications of these stereotypes, (3) have a sufficiently developed gender identity, and (4) have a well-formed conception of academic ability. If girls do not yet have the cognitive and social-cognitive abilities to possess these traits, it is unlikely that they will be impacted by stereotype threat.

The developmental timing of these requirements suggests that stereotype threat may begin to affect girls’ performance in the middle school years. Aronson and Good (2002) also argue that the particular conditions of the middle school climate combine with these developmental forces in order to create an environment in which stereotypes impact performance. It is at this age that social comparison is emphasized more in schools, and fitting in is a high priority for students (Harter, 1990, Harter, Whitesell, & Kowalski, 1992). Since stereotype threat effects are rooted in concern about how one’s performance is perceived by others, it seems possible that threat at this age may be particularly heightened. This timing also coincides with the age at which girls start to lose
confidence in their math abilities and soon after, they lose ground in math performance compared to boys, so it seems plausible that stereotypes begin to affect math performance at this age (Hyde, Fennema & Lamon 1990; Hyde, Fennema, Ryan, et al., 1990). Existing research examining stereotype threat effects in children has provided some evidence to support Good and Aronson’s (2002) age estimate, although other evidence suggests that effects may occur before girls reach middle school.

**Stereotype Threat Effects in Childhood and Adolescence: Empirical Findings**

There is mounting evidence that stereotype threat impacts performance prior to the time students reach college (Huguet & Regner, 2007; Muzzati & Agnoli, 2007). However, the current findings concerning the age at which stereotype threat begins to impact girls’ math performance are inconsistent. Some studies report evidence of stereotype threat effects with girls as young as kindergarten age, whereas others have not found these effects even in high school girls (Ambady, et al. 2001; Stricker & Ward, 2004). Table 1 displays some of the key design features and findings from the studies that have examined stereotype threat effects in children and adolescents.

Studies investigating stereotype threat effects in lower elementary school students have shown inconsistent results. Both Neuville and Croizet (2007) and Ambady et al. (2001) found stereotype threat effects for girls at this age, but Neuville and Croizet (2007) only found this effect on difficult mathematics items. In contrast, Muzzati and Agnoli (2007) did not find stereotype threat effects in second grade students and Neuville and Croizet (2007) actually found that girls in the stereotype threat condition did better than girls in the no-threat condition when given easy problems to solve.
There are only two studies that have been conducted with upper elementary school students and the findings of both students show little evidence of stereotype threat effects. Ambady et al. (2001) found no stereotype threat effects for girls in third through fifth grades, and actually found that girls this age performed better in the stereotype threat condition than in the no-threat condition. Across two experiments, Muzzatti & Agnoli (2007) found no evidence of stereotype threat effects for third, fourth, and fifth graders.

As discussed above, according to Aronson and Good (2002) one should expect that girls become susceptible to stereotype threat effects in middle and high school. There are several studies with middle schools students, though none involve a representative sample of American students. Three studies were done in Europe and one was done with Asian-American students. This research shows consistent evidence of stereotype threat effects in middle school students (Ambady et al., 2001; Huguet & Regner, 2007, 2009; Muzzatti & Agnoli, 2007 Experiment 2). For high school students, however, there appears to be some evidence of a stereotype threat effect in normal classroom conditions, but evidence for stereotype threat effects is not as clear in high-stakes standardized testing situations. In the regular classroom, Keller and Dauenheimer (2003) found stereotype threat effects in the general populations of ninth grade students whereas in Keller’s (2007) study stereotype threat led to poorer mathematics performance only for high school girls who were highly math-identified on difficult items.

Stricker and Ward (2004) examined stereotype threat in high school students taking the actual AP tests. The researchers varied whether students indicated their gender before (stereotype threat) or after (no-threat) they took the exam. They did not find
stereotype threat effects on girls’ math performance, even though the same manipulation has been shown to produce evidence of stereotype threat in college-age students. The authors believe these results indicate that stereotype threat effects do not impact performance in real-world testing situations. Some researchers argue that even though the findings were nonsignificant, they may still have practical significance (Danaher & Crandall, 2008). These researchers reanalyzed Stricker and Ward’s (2004) data and found that inquiring about gender before the test led 6% fewer girls to receive a score of 3 or higher. The authors stress that this leads to approximately 3,000 fewer girls receiving college credit for AP scores. Therefore, it can be argued whether stereotype threat effects, though not statistically significant, are practically important for high school students in real-world high-stakes testing environments.
### Table 1

*Summary of Literature on Stereotype Threat in Children and Adolescents*

<table>
<thead>
<tr>
<th>Article</th>
<th>Sample</th>
<th>Stereotype Threat Activation Method</th>
<th>Math Test</th>
<th>ST Effect?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambady, Shih, Kim &amp; Pittinsky, 2001</td>
<td>81 Asian-American girls grades K-8</td>
<td>Grades K-2: ST: draw picture of girl holding doll; No ST: draw landscape Grades 3-8: ST: questions related to gender; No ST: asked neutral questions</td>
<td>Iowa Test of Basic Skills for their age average ≈ 85%</td>
<td>Yes: K-2 and 6-8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No: Reverse in grades 3-5</td>
</tr>
<tr>
<td>Huguet &amp; Regner, 2007</td>
<td>504 French middle school students</td>
<td>ST: &quot;geometry test&quot; No ST: &quot;memory game&quot;</td>
<td>ReyOsterrieth Complex Figure: recall memory task of 2D line drawing average ≈ 60%</td>
<td>Yes</td>
</tr>
<tr>
<td>Huguet &amp; Regner, 2009</td>
<td>199 French middle school students</td>
<td>ST: ability in geometry No ST: ability in drawing</td>
<td>ReyOsterrieth Complex Figure average ≈ 53%</td>
<td>Yes</td>
</tr>
<tr>
<td>Keller, 2007</td>
<td>108 German high school students</td>
<td>ST: read that the test showed gender differences No ST: read that the test showed no gender differences</td>
<td>Items from TIMSS and math texts Difficult average ≈ 52% Easy average ≈ 93%</td>
<td>Yes: High Math ID on difficult items</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No: Low ID</td>
</tr>
<tr>
<td>Keller &amp; Dauenheimer, 2003</td>
<td>74 German high school students</td>
<td>ST: read that the test showed gender differences No ST: read that the test showed no gender differences</td>
<td>Items from TIMSS and math texts average ≈ 51%</td>
<td>Yes</td>
</tr>
<tr>
<td>Muzzatti &amp; Agnoli, 2007 (Exp. 1)</td>
<td>478 Italian elementary school children (2nd-5th grades)</td>
<td>ST: Saw picture of 9 male &amp; 1 female mathematician No ST: Saw picture of 9 flowers &amp;1 fruit Both: Students did math</td>
<td>Number subscale of Primary Mental Ability Test average ≈ 62%</td>
<td>No</td>
</tr>
</tbody>
</table>
problem based on picture

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>ST Description</th>
<th>No ST Description</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muzzatti &amp; Agnoli, 2007 (Exp. 2)</td>
<td>271 Italian 3rd, 5th, and 8th graders</td>
<td>ST: Vignette with 9 male and 1 female mathematician No ST: Parallel vignette with 9 flowers and 1 fruit Both: Students did math problem based on picture</td>
<td>Selected from standardized tests average ≈ 30%</td>
<td>Yes: 8th graders No: 3rd and 5th graders</td>
</tr>
<tr>
<td>Neuville &amp; Croizet, 2007</td>
<td>79 French 2nd grade children</td>
<td>ST: picture of a girl holding a doll No ST: landscape</td>
<td>Arithmetic problems Easy average ≈ 79% Difficult average ≈ 10%</td>
<td>Yes: difficult No: Reverse on easy</td>
</tr>
<tr>
<td>Stricker &amp; Ward, 2004 (Study 1)</td>
<td>1652 American students</td>
<td>ST: asked gender before No ST: asked gender after</td>
<td>AP Calculus AB test average ≈ 57%</td>
<td>No</td>
</tr>
<tr>
<td>Stricker &amp; Ward, 2004 (Study 2)</td>
<td>1341 incoming community college students</td>
<td>ST: asked gender before No ST: asked gender after</td>
<td>Community College Placement Test (elementary algebra and arithmetic) average ≈ 60%</td>
<td>No</td>
</tr>
</tbody>
</table>

These separate studies have employed very different methods and few studies have investigated multiple age groups with similar measures. However, there are two studies that have taken a developmental perspective. One of them is the study by Ambady et al. (2001), which examined stereotype threat in three age groups (lower elementary school, upper elementary school, middle school) for Asian-American girls. The findings were somewhat counterintuitive in that there were stereotype threat effects in lower elementary school students and middle school students, but girls in the upper elementary
years actually did better in the stereotype threat condition. These results must be interpreted with caution, as there are several weaknesses in the design of this study that make the results difficult to generalize, including the ease of the math test used as an outcome measure (average = 85%), the specific population, and the small sample size. In addition, stereotype threat was manipulated in different ways at different grades, making it difficult to compare stereotype threat effects across ages.

In the second study that took a developmental perspective, Muzzatti and Agnoli (2007) conducted two experiments examining the effects of stereotype threat across second to eighth grade Italian children (grades 2-5 in Experiment 1, and grades 3, 5, and 8 in Experiment 2). In Experiment 1, stereotype threat was activated by showing students a picture of 10 adult mathematicians: 9 male and 1 female. These investigators found no stereotype threat effects for students in the second through fifth grades. In Experiment 2 the researchers decided to use a more difficult math test and modified their stereotype threat activation. Along with the original manipulation, students answered a word problem with 9 boys and 1 girl as finalists in a math competition, which was more relevant to students than the adult mathematicians. In this experiment there was no stereotype threat effect found for third or fifth graders, but it was apparent in eighth graders. It was not clear if stereotype threat effects in eighth grade has to do with the fact that the participants were of a different age or that a different procedure was used.

As can be seen from the review of this literature, there is much inconsistency in the findings. Some of the inconsistencies may be due to the fact that these studies vary across a number of key dimensions. There was large variability in stereotype threat
activation methods, math performance measures, and experimental settings (i.e., high-stakes testing, classroom, or laboratory), making it difficult to compare results or draw inferences from findings across studies. Further, most of the available studies have been done either in Europe or in very specific population of American students and currently no developmental investigations have been done with American samples not confined to a particular ethnic group. In sum, the developmental research on stereotype threat effects is limited and the findings are mixed. The proposed dissertation is designed to deepen our understanding of this phenomenon by assessing stereotype threat effects in three age groups using parallel methods. This should allow us to delineate a developmental trajectory of stereotype threat effects on girls in mathematics.

**Exploring Mechanisms Underlying Stereotype Threat Effects**

Along with studying the occurrence of stereotype threat, it is important to understand through what mechanisms this phenomenon occurs so that psychologists and educators can develop policies and interventions targeting the processes involved. As is the case for gender differences, there is growing evidence suggesting that anxiety and working memory act as mediators of stereotype threat effects. Most notably, Schmader, Johns, and Forbes (2008) suggested a model where stereotype threat leads to anxiety because a woman fears that she will confirm the stereotype or be judged based on the stereotype. The anxiety is expressed as both components of anxiety, emotionality and worry. In this situation, women attempt to monitor and suppress these thoughts and feelings, which taxes their working memory resources. These working memory deficits,
in turn, lead to poorer math performance (Schmader et al., 2008). This mediation model is outlined in Figure 2.

Recently, there has been a growing body of literature examining these mediators. However, there are still large gaps in our understanding of how stereotype threat operates. Most of the existing studies looking into the nature of stereotype threat have focused on either emotional or cognitive components, with only a few studies testing both affective and cognitive processes potentially underlying the effects of stereotype threat. Further, there is only one study that identifies mediators of stereotype threat in adolescents and no work on children (Keller & Dauenheimer, 2003). This is a key area to study because it is possible that this phenomenon operates differently at younger ages. Below I review the available empirical evidence concerning the role of both anxiety and working memory in stereotype threat effects on math performance. It should be noted that some relevant bodies of work have been reviewed above (i.e., the relation between anxiety and working memory, the relation between anxiety and math performance, and the relation between working memory and math performance). Here I focus on studies that specifically examined mechanisms of stereotype threat effects as opposed to looking more generally at the factors underlying gender differences.
Figure 2

*Predicted Relation between Stereotype Threat and Math Performance through Anxiety and Working Memory*

**Stereotype threat and anxiety.** The question of whether inducing stereotype threat leads to increased anxiety has been studied extensively with inconsistent findings. These inconsistencies may be tied to two particular aspects in which the studies differ: the type of anxiety measured (emotionality or worry) and the timing of the measurement of anxiety.

Researchers who have investigated the emotionality component of anxiety generally find that stereotype threat does not lead to increased emotionality. These studies utilize self-report measures of emotionality, which ask participants to rate how much they are feeling particular emotions that represent the emotionality component of anxiety (e.g. anxious, jittery, nervous). A number of studies have found no evidence of a stereotype threat effect (Beilock et al. 2007; Brodish & Devine, 2009; Schmader, 2002; Schmader & Johns, 2003; Spencer et al., 1999). One study, however, did find that emotionality was increased under stereotype threat, specifically when using a scale that measured a broad range of emotions – making it less obvious that anxiety was under
investigation (Marx & Stapel, 2006). Overall, findings suggest that this aspect of anxiety may not be affected by stereotype threat, unless the goal of the measure is less obvious.

Less work has examined the worry component of anxiety, but these studies have found evidence that stereotype threat leads to worry and some studies have suggested that worry also acts as a mediator of stereotype threat effects on math performance. Cadinu and colleagues (2005) asked women to report their thoughts during testing and found that stereotype threat led to more worry-related intrusive thoughts. In another study, Beilock et al. (2007) asked participants after the test to retrospectively report how much they had worried during the test and found that stereotype threat led to worry. Additionally, Brodish & Devine (2009) found that women under threat reported more worry on a self-report measure. Both Brodish and Devine (2009) and Cadinu and colleagues (2005) also found that worry mediated the effect of stereotype threat on mathematics performance. Taken together, the research findings on emotionality and worry show limited evidence for the role of emotionality in stereotype threat effects but solid evidence that worry may be involved.

Another variable that has the potential to impact the findings of stereotype threat effects on anxiety is the timing of the administration of the anxiety measure. In the literature on stereotype threat effects, anxiety measures have been given before, during, or after the math test. It seems critical in order to make an argument that anxiety is a mediator of the effect of stereotype threat on mathematics performance, that anxiety be measured after the stereotype threat manipulation and before the math assessment (Spencer et al., 1999). It also seems likely that anxiety occurs throughout the entire
testing situation, such that it may be accurately measured at multiple times during the test. The findings of Cadinu and colleagues (2005) suggest that measuring anxiety at multiple times during testing may be a useful way of understanding its relation with stereotype threat effects as there was evidence of increased anxiety throughout the stereotype threatened situation. Studies that have measured anxiety after mathematics performance tend to not find stereotype threat effects on anxiety (Keller & Dauenheimer, 2003; Marx & Stapel, 2006; Schmader & Johns, 2003). Based on this research, it is clear that assessing anxiety both before and during the math assessment may be useful in fully understanding the anxiety occurring during math testing situations.

**Stereotype threat, anxiety, working memory and math performance.** There is a growing body of literature suggesting that stereotype threat impacts working memory performance and some evidence that working memory acts as a mediator of the effect of stereotype threat on math performance. However, there is mixed evidence for the role of anxiety in this relationship. Beilock and colleagues (2007) found that worry mediated the negative impact of stereotype threat on verbal working memory. Schmader and Johns (2003) found that verbal working memory mediated the effect of stereotype threat on math performance, but there was no evidence that the emotionality component (as measured by self-report after the test) was involved in this relation.

Recently, Johns, Inzlicht, and Schmader (2008) conducted a series of experiments to test anxiety and working memory as mechanisms through which stereotype threat effects on performance occur and found indirect evidence for this model. They found that women under stereotype threat showed signs that they were suppressing their anxiety and
that this anxiety suppression led to poorer working memory performance. In addition, stereotype threat effects were partially explained by working memory performance. These results suggest that stereotype threat effects can be explained by this mediating chain of relations. However, this study only indirectly assessed anxiety as a mediator and the measure used to show anxiety suppression lends itself to alternative explanations, including the possibility that these women may have actually been less anxious. Overall, the findings of research that investigates the entire model shows that the worry component of anxiety or anxiety suppression may be related to stereotype threat, working memory, and math performance, but perhaps not emotionality, at least when an scale consisting of only items assessing emotionality is used after testing. In addition, these studies exclusively focused on verbal working memory, thus the role of visual working memory is unknown.

In summary, researchers have found evidence for multiple relations involved in the proposed model positing that stereotype threat leads to anxiety, which taxes working memory, thus leading to poorer mathematics performance. Still, there is little systematic work involving all of the components and exploring both components of anxiety (emotionality and worry) as well as two types of working memory (verbal and visual). Critically, all of the existing work testing the role of anxiety and working memory in stereotype threat has been done with adults and it remains to be determined whether these factors can explain stereotype threat effects on children and adolescents’ math performance.
Present Research

Thus, the current literature has a number of areas in which more research is needed. First, there is a lack of a developmental perspective with regard to the study of gender differences and stereotype threat effects, and few studies address these questions using parallel measures across multiple age groups. Second, no studies have examined the entire explanatory chains simultaneously – looking at the ability of anxiety and working memory to explain gender differences and stereotype threat effects. Further, given the inconsistencies in findings with respect to the different components of anxiety and different types of working memory, it is important to systematically examine these variables in relation to the same measures of math performance. In the present dissertation, these issues have been addressed in two studies.

Study 1. This study examined the nature of gender differences and stereotype threat effects as well as the role of anxiety in these relations across school-age children and adolescents. Thus, in exploring potential mediators of the relation between gender and math performance, Study 1 focused on just the anxiety piece of the model to test this relation prior to assessing whether working memory is also involved. There were four specific goals. The first two goals were to examine (1) gender differences in math performance across development and (2) stereotype threat effects across development. To address these goals, high performing students in the fourth, eighth, and twelfth grades were assigned to the stereotype threat condition or the no-threat condition and given a difficult mathematics test. Difficult items were chosen along with a high performing sample of students because it has been shown in the literature that both gender
differences and stereotype threat effects are more often found on difficult measures with high performing students (Gibbs, 2010; Hyde, Fennema, & Lamon, 1990; Neuville & Croizet, 2007; Steele, 1997). Students were chosen across a large span of development in order to assess how gender differences and stereotype threat effects might vary across ages. Research suggests that gender differences and the impact of stereotype threat effects may not emerge until middle to high school (Aronson & Good, 2002; Lindberg et al., 2010), but some work suggests that both gender differences and stereotype threat effects may occur earlier in development on more difficult tasks (Gibbs, 2010; Neuville & Croizet, 2007).

The other two goals of this study involved delving further into the underlying mechanisms related to both gender differences and stereotype threat effects. The third goal was to examine, if there are gender differences on the math test, whether anxiety acts as a mediator of the relation between gender and math performance across development. The fourth goal was to explore whether anxiety acts as a mediator of stereotype threat effects across development, if these effects are found. Because stereotype threat is a situational phenomenon, where the particular demands of the testing situation may activate threat, it made more sense to look at state anxiety rather than overall trait anxiety in trying to understand the underlying mechanisms for both gender differences and stereotype threat effects. In order to understand the role of the emotionality and worry components of state anxiety, both of these components were measured. In addition, based on research that has shown that gathering information about anxiety both before and during the testing situation can be useful, this study assessed
anxiety at two time points (Cadinu et al., 2005). This study contributes to the literature by examining this full relation, instead of bivariate relations between constructs. In addition, it takes a developmental perspective by investigating these relations across multiple ages with parallel methods.

**Study 2.** Study 2 further investigates the relations assessed in Study 1. Whereas the first study focused specifically on anxiety, the second study looked at potential cognitive mechanisms underlying anxiety’s effect on performance in the context of gender and stereotype threat. If anxiety is found to be related to math performance, this study will examine whether working memory mediates this relation. Due to the inconsistent findings in the literature in regard to the type of working memory that is affected by anxiety and related to math performance, this study assesses both verbal and visual working memory (Gathercole, et al., 2004; Holmes, et al., 2008). The specific details of the methodology of Study 2 are largely informed by the findings of Study 1, which provide the foundational information about the pattern of relation between particular measures of anxiety and math performance at different ages.
Chapter 3: Study 1

In this experimental study, participants were randomly assigned to stereotype threat and no-threat conditions and given anxiety measures and a difficult math test. The stereotype about boys being better than girls at math was activated in the stereotype threat condition, but not in the no-threat condition; in all other respects, the two experimental conditions were equivalent. The study employed a 2 (condition: stereotype threat, no-threat) x 2 (gender: girls, boys) x 3 (grade: fourth grade, eighth grade, twelfth grade) design. In addition, mediation models were examined with emotionality and worry as mediators of both the gender difference in math and stereotype threat effects on math performance. These models were tested with statistical mediation analyses, through a series of regression analyses (Dearing & Hamilton, 2006; Hayes, Preacher, & Myers, 2011). Study 1 focused on four research questions:

1) Are there gender differences in math test performance across fourth, eighth and twelfth grades?

It is expected that boys will have higher performance than girls on the math test at the eighth and twelfth grade levels. The prediction concerning twelfth graders is based on research that suggests that gender differences emerge in high school (Hyde, Fennema & Lamon, 1990; Lindberg et al., 2010). Researchers have also shown that gender differences in high ability samples emerge by middle school (Entwisle et al., 1994). Thus, since the sample in this study is high performing, gender differences may emerge by the eighth grade. With respect to fourth grade students, there is no definitive prediction. On one hand, most studies find no gender differences in math among contemporary students
at this age (Hyde, Fennema, & Lamon, 1990; Lindberg et al., 2010). On the other hand, some investigators suggest that the complexity of the test may override this developmental pattern (Gibbs, 2010). Thus gender differences may emerge at an earlier age in this study because the math assessments are difficult and the sample is high achieving (Gibbs 2010).

2) Does stereotype threat with regard to females and math impact the math test performance of school-age students? Are there differences in the effects of math stereotype threat across development?

It is expected that stereotype threat will impact girls’, but not boys’, performance at the eighth and twelfth grade levels but not in the fourth grade. Thus, at the two older grades it is expected that girls in the stereotype threat condition will perform more poorly compared to girls in the no-threat condition and that boys will not differ in their performance across conditions. This prediction is based on research by Good and Aronson (2002), which suggests that girls will not be affected by the gender stereotype threat concerning math until around the age of 11 or 12. It is at this time that the developmental requirements for stereotype threat (such as awareness of gender stereotyping and understanding the implications of negative stereotypes) are in place. Thus, if these traits develop at this time, and are related to stereotype susceptibility, it is likely that students at the fourth grade level – who are approximately 9 or 10 years old, will not yet be susceptible to stereotype threat effects. However, students in eighth and twelfth grades have likely met these developmental requirements and will therefore be more likely to be impacted by stereotype threat.
3) If there is a gender difference in math performance, do the two components of state anxiety (emotionality and worry) mediate this relation? Does this mediation relation differ as a function of grade level?

It is proposed that the relation between gender and math performance will be mediated by both the emotionality and worry components of state anxiety, but that the worry component may act as a stronger mediator (Deffenbacher, 1980; Hembree, 1988). Current research suggests that gender differences in the worry and emotionality components of anxiety begin to develop during the elementary school age, but become more pronounced in middle and high school (Hembree, 1988). It is also around elementary school that anxiety begins to impact performance (Hembree, 1988). Therefore, if a gender difference is found at all three grade levels, it is expected that both components of anxiety will mediate the relation in all grades, but they may be stronger mediators at older ages (Hembree, 1988).

4) If there are stereotype threat effects on girls’ math performance, do the two components of state anxiety (emotionality and worry) mediate this relation? Does this mediation relation differ as a function of grade level?

It is predicted that stereotype threat effects will be mediated by at least the worry component of state anxiety such that stereotype threat leads to heightened worry in girls, which leads to poorer math performance. Assuming stereotype threat only impacts girls, as predicted, this analysis will only include girls. Thus the girls in the stereotype threat condition will be compared to girls in the no-threat condition. Based on past research, it is expected that stereotype threat will lead to higher levels of worry as measured by both
a thought listing measure and a self-report measure (Beilock, et al, 2007; Brodish & Devine, 2009; Cadinu et al., 2005; Spencer et al., 1999). With regard to the emotionality component, it is unclear whether it will be increased in stereotype threat conditions. Most research evidence suggests that stereotype threat does not lead to increased emotionality, but one study, using a less obvious anxiety measure, did find stereotype threat effects on emotionality (Marx & Stapel, 2006). Research shows that both components of anxiety are related to math performance, and that the worry component of anxiety is more related to performance than the emotionality component (Hembree, 1988). Based on these research findings, it is likely that the worry component of state test anxiety may mediate the relation between stereotype threat and math performance, but it is unclear whether the emotionality component will also serve as a mediator.

Method

Participants. Fourth, eighth, and twelfth grade participants were recruited from five high performing suburban schools in the Boston area. Table 2 presents demographic and socioeconomic information about participating schools (Massachusetts Department of Education, 2010). Table 3 presents information about the math performance of students from participating schools, based on state standardized test scores (Massachusetts Comprehensive Assessment System (MCAS), Massachusetts Department of Education, 2010).
Table 2

Demographic Information, by School

<table>
<thead>
<tr>
<th>Income Level (%)</th>
<th>Race (%)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low-income</td>
<td>White</td>
<td>Asian-American</td>
<td>Hispanic</td>
<td>African-American</td>
<td>Other</td>
</tr>
<tr>
<td>Fourth Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School 1</td>
<td>11</td>
<td>73</td>
<td>16</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>School 2</td>
<td>18</td>
<td>73</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Eighth Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School 3</td>
<td>13</td>
<td>80</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Twelfth Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School 4</td>
<td>4</td>
<td>89</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>School 5</td>
<td>8</td>
<td>81</td>
<td>12</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 3

*Percentage of Students at Each Achievement Level on State Math Test, by School*

<table>
<thead>
<tr>
<th></th>
<th>Advanced</th>
<th>Proficient</th>
<th>Needs Improvement</th>
<th>Failing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fourth Grade</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State</td>
<td>20</td>
<td>40</td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td>School 1</td>
<td>40</td>
<td>36</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>School 2</td>
<td>50</td>
<td>35</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td><strong>Eighth Grade</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State</td>
<td>16</td>
<td>33</td>
<td>30</td>
<td>21</td>
</tr>
<tr>
<td>School 3</td>
<td>24</td>
<td>40</td>
<td>29</td>
<td>7</td>
</tr>
<tr>
<td><strong>Twelfth Grade</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State</td>
<td>43</td>
<td>29</td>
<td>19</td>
<td>9</td>
</tr>
<tr>
<td>School 4</td>
<td>71</td>
<td>21</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>School 5</td>
<td>74</td>
<td>21</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note.* The numbers in the Table represent students’ scores from the last time they took the MCAS test. Third grade scores from 2009 were used for fourth graders, seventh grade scores from 2009 were reported for eighth graders, and tenth grade scores from 2008 were reported for twelfth graders.
As seen in Table 3, only schools where students performed substantially better than the state average in the target grade on the Math MCAS test were recruited. The decision to include higher performing students was based on past research that has shown that (1) gender differences are more pronounced and develop earlier in high-achieving students (Entwisle et al., 1994; Lindberg et al., 2010) and (2) that stereotype threat effects are stronger in these students because they are more likely to identify with the academic domain of mathematics (Steele, 1997).

Participants were 71 fourth graders (41 boys, 30 girls), 107 eighth graders (40 boys, 67 girls), and 147 twelfth graders (70 boys, 77 girls). To recruit participants, school principals were initially contacted via email. The decision to take part in the study was made by the principals in consultation with the teachers in the target grade levels. For students under the age of 18, parent consent as well as student assent was obtained. For students over 18, student consent was obtained.

Materials. During the study, stereotype threat was experimentally manipulated, and three state anxiety measures and a math test were administered to participants. The anxiety measures were adapted from adult measures; items were sometimes edited to be developmentally appropriate for both children and adolescents.

Stereotype threat manipulation. Students in both conditions were presented with a sample math problem in order to manipulate stereotype threat. In the stereotype threat condition, the sample problem portrayed a situation in which a much larger proportion of boys than girls received a math award or were chosen for the math team based on their performance on a math test. In the no-threat condition, students were presented with a
sample problem about a topic unrelated to gender or math (groups of students attending a field trip). Students then chose the correct answer to the math problem from among five possible choices. This procedure (having the stereotype threat manipulation as a sample word problem) was chosen because it makes it more likely that the students fully process the gender-related contents of the word problem as they solve that problem. Further, the manipulation of stereotype threat used in this procedure is less obvious than that used in many other studies, for example, when participants are explicitly told that girls are worse than boys in math, which is very unlikely to occur in real life. The mathematical knowledge required for the math problem was different for each grade level, so as to make it age-appropriate. At each grade level, the computational task required to solve the math problem was identical across the two experimental conditions (Muzzatti & Agnoli, 2007).

For example, the eighth graders in the stereotype threat condition read the sample word problem stating:

At the Miller Middle School, the boys were much better at math than the girls. The math teachers chose the 20 students with the highest math test scores for the math team to represent the school at the statewide math competition. Eighteen of the students were boys and two were girls. What proportion of the students on the math team were boys?

The eighth graders in the no-threat condition read:

At the Miller Middle School, students were invited to participate in a special field trip, but there were only 20 spots available. The teachers chose 18 students from
Ms. Fletcher’s homeroom and two other students from Ms. Johnson’s homeroom.

What proportion of the students going on the field trip were from Ms. Fletcher’s homeroom?

The sample word problems and answer choices for both conditions for fourth, eighth and twelfth grades are provided in Appendices A, B, and C, respectively.

State anxiety. Students were given three measures of state anxiety designed to measure both the emotionality and worry components.

Worry was measured using a thought listing worry measure (hereafter referred to as thought listing measure) and a self-report worry scale (hereafter referred to as worry scale). For the thought listing measure, students were asked to “Please write three things that you are feeling or thinking about right now.” This measure was designed to capture a broad range of affective states. Students’ comments were coded based on their content. The coding categories were: negative math-related thoughts (i.e., “I am not good at math”), hate for math (i.e., “I have always hated math”), generic distress (i.e., “I am tired”), self-confidence (i.e., “I am really good”), neutral reference to the test (i.e., “Probably one of these alternatives are correct”), not knowing what to write (i.e., “Nothing”), guess (i.e. “For this one I had no clue”), and other (i.e., “Pizza”). Thoughts that were coded as negative math-related thoughts were analyzed as a measure of state anxiety (Cadinu, et al., 2005). This measure has been reliably used to measure anxiety during the entire testing situation, therefore, students were asked to report their thoughts both before and in the middle of the math test.
The worry scale was made up of four questions (see Appendix D). Participants rated their feelings on a scale from 1 (strongly disagree) to 4 (strongly agree) for statements about their worry (e.g., I am worried that I may not do well on this test.), two of which were reverse coded (e.g., I feel very confident about my performance on the test I’m about to take.) (adapted from Morris, Davis, & Hutchings, 1981). A worry score was calculated by summing the responses on the four items (after reverse coding).

Emotionality was measured using an emotion list (adapted from Marx & Stapel, 2006). Participants saw a list of 16 emotions and rated the extent to which they were feeling each emotion on a scale from 1 (not at all) to 4 (very much). This emotion list included five items that measured the emotionality component of anxiety (afraid, anxious, confident (reverse coded), nervous, uncertain) and 11 filler items (full scale is presented in Appendix E). A total score for emotionality was calculated by summing the responses for each of the five anxiety items. Filler items were included so that it would not be as obvious that the measure was focused on anxiety.

Math performance. Students answered twelve age-appropriate multiple-choice mathematics items (six Number/Algebra items, six Geometry/Measurement items) sampled from NAEP, TIMSS, and MCAS mathematics assessments for fourth, eighth and twelfth grades (see Appendices F, G, and H). These items were selected to represent a high level of difficulty because gender differences are often found on more difficult items and stereotype threat effects tend to impact performance most on difficult tests (Gibbs, 2010; Keller, 2007; Neuville & Croizet, 2007; Steele, 1997). Only NAEP and TIMSS items with accuracy rates lower than 40% in national samples and MCAS items
with state accuracy rates below 60% were selected. At each grade level, Number/Algebra and Geometry/Measurement items had approximately equal difficulty level (between 30% and 35% in national and state samples) averaged across items in each set. Students completed the six Number/Algebra problems in one block and the six Geometry/Measurement problems in another. The order of the blocks was counterbalanced across children within each experimental condition. Fourth and eighth graders were given five minutes to complete each section and twelfth graders were given six minutes for each section.

Since the test was timed, statistical analyses for math performance could be done in two different ways. One involves the proportion of items answered correctly on the math test (out of 12). This method penalizes students for working slowly and not completing the test by assigning them a 0 for items that they did not complete. Therefore, this measure is more related to the processing efficiency theory, or how long it takes one to complete the items (Osborne, 2006). As this theory states, if students’ working memory is heavily taxed, it takes them longer to complete items, which decreases the number of items that they can complete. A second way to analyze math performance is to examine the relative accuracy of performance on the math test (the number of items answered correctly divided by the number of items attempted). This analysis assesses how well students perform on the items that they have time to complete, without penalizing them for not finishing the test. In the present study, students’ math performance was coded in both ways.
Procedure. Students were separated into two groups based on the condition they were assigned to, and each group was tested in a separate room. Two researchers conducted the study, alternating between conditions – e.g., about half the students in the stereotype threat condition were tested by one researcher and the other half by the other researcher. Instructions were read aloud for students at all grade levels and students were encouraged to ask questions if they were unsure of what to do. For all grades, the sample math problem was read by the researcher as the participants read along. Also, for fourth grade students, the items from the anxiety measures were read aloud to minimize the possible confound of reading ability with survey responses.

Figure 3 provides an outline of the steps involved in the Study 1 procedure. During Step 1 students read the sample math problem that either activated stereotype threat or did not. They were required to answer the math question about the contents of the word problem. In Step 2 students completed the three measures of state anxiety: the thought listing measure, the emotion list, and the worry scale. In Step 3, students were reminded of the sample math problem they had done, to reinforce the stereotype threat manipulation for those in the stereotype threat condition. At this point, the students were also told about the remaining math items. In the no-threat condition, they were told that they were going to do some math problems, whereas in the stereotype threat condition students were told they would be taking a math test. This was done in order to make the math assessment sound more evaluative in the stereotype threat condition, which has been shown to increase stereotype threat effects (Aronson & Steele, 2005; Good & Aronson, 2008; Steele, 1997). In Step 4 students completed the first half of the
mathematics test for their grade level (either the Number/Algebra section or the Geometry/Measurement section). During Step 5 they completed the thought listing worry measure only. In Step 6 they completed the second half of the mathematics test. The entire testing session took approximately 30-45 minutes.

Figure 3

*Study 1 Procedure*

Results

**Preliminary analyses.** In the preliminary analyses, the data were inspected to examine normality, test for outliers, and measure reliability.

**Normality check.** The examination of the shape of the distribution for the variables at each grade level showed that several distributions were positively skewed. This conclusion was based on dividing the skewness statistic by its standard error and
comparing this z-score to the critical z-score of 1.96 (Glass & Hopkins, 1996). Scores on the emotion list were skewed for all three grade levels and scores on the worry scale were skewed for fourth graders. Natural log transformations were conducted for these variables, which significantly lowered the skew statistics to acceptable levels. These transformed variables were used in subsequent statistical analyses.

**Outliers.** Outliers were identified based on both the sample distribution for each variable and the residual statistics for the regression analyses used for the mediation analyses at each grade level. In regard to the sample distribution, outliers were identified based on the interquartile range (IQR, the middle 50% of the distribution). Cases that fell outside of the 1.5 x IQR range were identified as potential outliers (Glass & Hopkins, 1996). Then, the residuals of the regression analyses run for the mediation analyses were inspected to identify cases that were outside of the recommended cutoff point \( \frac{3}{\sqrt{n}} \) for Standardized DFBeta (Pedhazur, 1997). As a result, 7 cases were removed (3 from fourth grade, 2 from eighth grade, 2 from twelfth grade) because they were outside of the cutoffs on both of these outlier measures. Thus, all subsequent statistical analyses were run on the remaining sample of 68 fourth graders (39 boys, 29 girls), 105 eighth graders (40 boys, 65 girls), and 145 twelfth graders (69 boys, 76 girls).

**Reliability.** The reliabilities for the measures of state anxiety were high in most cases. For the thought listing measure, interrater reliability was calculated based on the coding of two independent raters. Agreement occurred 93% of the time. For all of the other scales, Cronbach’s alphas were computed as a measure of internal consistency (Table 4). For the worry scale, the reliabilities were high at all three grade levels. The
ratings of the anxiety items from the emotion list, which measured emotionality, hung together well in both the eighth and twelfth grades ($\alpha = .741$ and .742, respectively) but had lower reliability in the fourth grade ($\alpha = .530$). Fourth graders’ responses to the item “anxious” were not highly correlated with their responses on other items, likely because students at this age may not know the definition of the word anxious. To examine the effect of this item on the overall measure, analyses were conducted both with the full version of the scale and with a more reliable version of the scale ($\alpha = .611$) that excluded the item anxious. The results of analyses with “anxious” removed were similar to those when the entire scale was used. Therefore, in order to use parallel measures across grades, the analyses with the full scale of five items are reported.

The reliabilities for the math test were moderate at each grade level. Since the test was specifically designed to measure a broad range of concepts in both Number/Algebra and Geometry/Measurement, a high reliability would not be expected. The lower reliabilities at the higher grade levels are likely a reflection of the fact that there are more diverse concepts covered in mathematics at these grade levels. However, these reliabilities are still acceptable, especially given the relatively small number of items included in each test (Hair, Black, Babin & Anderson, 2010).
Table 4

*Reliabilities for Measures by Grade*

<table>
<thead>
<tr>
<th></th>
<th>Fourth Grade</th>
<th>Eighth Grade</th>
<th>Twelfth Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotionality</td>
<td>.530</td>
<td>.741</td>
<td>.742</td>
</tr>
<tr>
<td>Worry Scale</td>
<td>.885</td>
<td>.846</td>
<td>.781</td>
</tr>
<tr>
<td>Math Test</td>
<td>.800</td>
<td>.668</td>
<td>.639</td>
</tr>
</tbody>
</table>

**Descriptive statistics.**

*Means and standard deviations.* The examination of frequencies of different types of responses on the thought listing measure showed that a very small portion of students indicated that they were having any negative math-related thoughts either before the test (fourth grade: 15%; eighth grade: 12%; twelfth grade: 7%) or during the test (fourth grade: 14%; eighth grade: 12%; twelfth grade: 8%). Because of this floor effect, the thought listing measure was not used in statistical analyses and the worry scale was used as the only measure of worry. Table 5 shows descriptive statistics for the remaining measures.
Table 5

*Descriptive Statistics (Means and Standard Deviations) by Grade, Gender, and Stereotype Threat Condition*

<table>
<thead>
<tr>
<th></th>
<th>Girls</th>
<th></th>
<th>Boys</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stereotype Threat</td>
<td>No-Threat</td>
<td>Stereotype Threat</td>
<td>No-Threat</td>
</tr>
<tr>
<td>Fourth Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotionality</td>
<td>8.57 (3.01)</td>
<td>7.33 (2.50)</td>
<td>6.44 (1.20)</td>
<td>8.62 (2.33)</td>
</tr>
<tr>
<td>Worry</td>
<td>7.29 (2.13)</td>
<td>6.53 (2.33)</td>
<td>4.72 (1.07)</td>
<td>7.14 (2.99)</td>
</tr>
<tr>
<td>Math Score</td>
<td>0.61 (0.30)</td>
<td>0.56 (0.29)</td>
<td>0.75 (0.18)</td>
<td>0.69 (0.25)</td>
</tr>
<tr>
<td>Eighth Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotionality</td>
<td>8.52 (2.98)</td>
<td>9.55 (3.11)</td>
<td>6.44 (2.20)</td>
<td>6.91 (2.31)</td>
</tr>
<tr>
<td>Worry</td>
<td>8.31 (2.12)</td>
<td>8.24 (2.28)</td>
<td>6.94 (2.98)</td>
<td>5.91 (1.48)</td>
</tr>
<tr>
<td>Math Score</td>
<td>0.58 (0.22)</td>
<td>0.55 (0.20)</td>
<td>0.67 (0.24)</td>
<td>0.68 (0.20)</td>
</tr>
<tr>
<td>Twelfth Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotionality</td>
<td>8.17 (2.20)</td>
<td>8.15 (3.07)</td>
<td>7.18 (2.71)</td>
<td>6.59 (1.72)</td>
</tr>
<tr>
<td>Worry</td>
<td>7.53 (1.87)</td>
<td>7.85 (2.29)</td>
<td>6.40 (2.05)</td>
<td>6.28 (2.02)</td>
</tr>
<tr>
<td>Math Score</td>
<td>0.42 (0.20)</td>
<td>0.48 (0.24)</td>
<td>0.57 (0.19)</td>
<td>0.57 (0.22)</td>
</tr>
</tbody>
</table>

*Note.* Scale for Emotionality measure is from 5 to 20; Scale for Worry scale is from 4 to 16; Scale for math score is proportion correct.
Correlations. Tables 6, 7, and 8 show bivariate correlations for fourth, eighth, and twelfth graders, respectively. These correlations show that at the eighth and twelfth grade levels, all of the variables are significantly related to each other in expected directions.

There are gender differences favoring boys in worry, emotionality, and math performance and worry and emotionality are both significantly negatively correlated with math performance. Worry and emotionality are also significantly correlated with each other. At the fourth grade, there are fewer correlations between the measures. There is a gender difference in math performance, but no gender differences in emotionality or worry. Worry and emotionality are related to each other, but neither measure is related to math performance.

Table 6

Correlations among Measures, Fourth Grade

<table>
<thead>
<tr>
<th></th>
<th>Gender</th>
<th>Emotionality</th>
<th>Worry</th>
<th>Math Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotionality</td>
<td>-.065</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worry</td>
<td>-.176</td>
<td>.565**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Math Score</td>
<td>.255*</td>
<td>-.117</td>
<td>-.107</td>
<td>1</td>
</tr>
</tbody>
</table>
### Table 7

*Correlations among Measures, Eighth Grade*

<table>
<thead>
<tr>
<th></th>
<th>Gender</th>
<th>Emotionality</th>
<th>Worry</th>
<th>Math Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td>-.381**</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Emotionality</td>
<td>-.386**</td>
<td>.599**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Worry</td>
<td></td>
<td></td>
<td>.241*</td>
<td></td>
</tr>
<tr>
<td>Math Score</td>
<td></td>
<td>-.387**</td>
<td>-.404**</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table 8

*Correlations among Measures, Twelfth Grade*

<table>
<thead>
<tr>
<th></th>
<th>Gender</th>
<th>Emotionality</th>
<th>Worry</th>
<th>Math Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td>-.238**</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Emotionality</td>
<td>-.313**</td>
<td>.504**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Worry</td>
<td></td>
<td></td>
<td>.263**</td>
<td></td>
</tr>
<tr>
<td>Math Score</td>
<td></td>
<td>-.229**</td>
<td>-.318**</td>
<td>1</td>
</tr>
</tbody>
</table>

**Main statistical analyses.** The results were analyzed in several steps. First, I examined students’ math performance across groups; specifically comparing boys and girls across experimental groups, and across different grade levels. If gender differences or stereotype threat effects existed, further analyses were done to determine whether the
components of anxiety mediated the relation between either gender and math performance or stereotype threat and math performance.

**Gender differences and stereotype threat effects across development.** In order to investigate gender differences in math performance as well as the effects of stereotype threat, a factorial analysis of variance (ANOVA) was run. This analysis aimed at addressing the first two research questions of Study 1. There were three between-subject variables: condition (stereotype threat, no-threat), gender (girls, boys), and grade level (fourth grade, eighth grade, twelfth grade). The dependent variable was the proportion of items answered correctly (out of all twelve items on the math test). It should be noted that for all analyses, results were similar when using the proportion correct out of all twelve items and the proportion correct out of items attempted. For ease of presentation, only the results of analyses using proportion correct out of all twelve items are reported.

My prediction in regard to gender differences was that there would be significant differences in math test performance for eighth and twelfth graders. With respect to fourth graders, there was not a clear prediction, as some researchers suggest that gender differences emerge in middle and high school but others argue that these differences could emerge earlier on challenging tests. If the gender difference in fourth graders is nonexistent or smaller than that for older students, there would be a significant two-way interaction between gender and grade indicating that gender differences are not the same across grade levels. If there is a comparable gender differences across grades this would be indicated by a main effect of gender and no interaction between gender and grade.
My prediction with regard to stereotype threat was that stereotype threat would impact girls’ math performance at the eighth and twelfth grade levels and not at fourth grade. This prediction would be supported if there were a significant three-way interaction between stereotype threat condition, gender, and grade indicating that stereotype threat effects differ across grade level. Then, simple effects tests would show that the interaction between gender and condition was significant only at the eighth and twelfth grade levels and these interactions would indicate that stereotype threat detrimentally impacts girls’ math performance (and not boys) at these ages.

When running ANOVAs, unequal sample size can sometimes be a problem. However, it is only an issue if there is heterogeneity of variance (Keppel & Wickens, 2004). Levene’s test of homogeneity of variance was nonsignificant, indicating that there is no evidence of heterogeneity and thus the unequal sample sizes are not a problem with subsequent analyses (Keppel & Wickens, 2004).

The ANOVA results revealed main effects of both grade ($F(2, 306) = 12.05, p < .01$) and gender ($F(1, 306) = 19.89, p < .01$), but not condition, and no significant interactions among any of the variables (See Figure 4). In order to interpret the main effect of grade, LSD posthoc comparisons between the three grade levels were run. These results showed that students in the twelfth grade had significantly lower performance on the math test than students in the fourth and eighth grades (fourth $M = 0.66$, eighth $M = 0.61$, twelfth $M = 0.51$). Since students in different grades took different tests, this main effect is not meaningful, as it likely reflects slight differences in the difficulty levels of the tests. With respect to gender, the existence of a main effect and no interaction with
grade indicates that boys performed better than girls across grade levels (Boys $M = 0.64$, Girls $M = 0.52$). An examination of effect sizes at each grade level shows that gender differences were similar at all three grades (fourth $d = 0.59$; eighth $d = 0.51$; twelfth $d = 0.54$). Thus, the present study confirmed the prediction that gender differences in math performance exist at the eighth and twelfth grades and indicated that gender differences on a challenging test also exist at the fourth grade level. However, the results showed no evidence for stereotype threat effects at any grade level, in contrast to the prediction that these effects would exist at the eighth and twelfth grades.

Figure 4

*Math Test Performance Across Gender and Grade*
Anxiety as a mediator of gender differences in math performance. In order to address research question 3, whether the relation between gender and math test performance could be explained by emotionality and/or worry, a mediation analysis was conducted at each grade level (See Figure 5).

Analytic strategy. To test for mediation, a series of regression models were run. The first three regression analyses tested the initial requirements for mediation – that the predictor (gender) is related to both the outcome measure (math performance) and the potential mediators (emotionality and worry). As a next step in examining mediation, a regression model that included gender, emotionality, and worry as predictors of math test performance was run. Mediation is likely evident if the relation between gender and math test performance disappears or significantly decreases after the addition of emotionality and worry into the model. Note that stereotype threat condition was included as a covariate in the analyses in order to remove any variance accounted for by the differences in experimental condition.
To complete the mediation analysis, I used bias-corrected bootstrapping to estimate confidence intervals for the mediated effect (Dearing & Hamilton, 2006; Preacher & Hayes, 2008). This method has been recommended over the causal steps approach by a number of researchers (Dearing & Hamilton, 2006; Preacher & Hayes, 2004; Preacher & Hayes, 2008). This strategy involved running a procedure that selected 5,000 bootstrap samples with replacement from the current sample and estimated the regression coefficients within each of these bootstrap samples (Preacher & Hayes, 2008). The bootstrapping algorithm produced an average of these estimates, which was used in interpreting the mediation results. The analyses were run with the INDIRECT SPSS macro created by Preacher and Hayes.
This method allowed me to examine the overall indirect effect of gender through anxiety, including both the emotionality and worry components, as well as the specific indirect effects for each mediator separately. In order to estimate these indirect effects, confidence intervals were calculated for the product of the path coefficients \((ab)\) between the predictor and mediator(s) \((a)\) and between the mediator(s) and the outcome \((b)\). For the overall indirect effect, the product of the path coefficients is equivalent to the difference between \(c\) (the total effect of gender; the coefficient for the simple regression analysis with gender and math performance) and \(c'\) (the direct effect of gender; the coefficient based on the regression of math performance after the anxiety measures are included). If the product of the path coefficients for an anxiety component is significant, this indicates that there is a significant indirect effect of gender on math performance through that component. In order to determine significance, the 95% confidence intervals for each of the state anxiety components were examined. If the confidence interval overlapped with 0, then the indirect effect was not significant, however, if the confidence interval did not overlap with 0, the indirect effect was considered significant (Preacher & Hayes, 2004).

*Estimation of the mediation model.* First, I tested the initial requirements for mediation by running three regression analyses at each grade level to test whether gender was related to both the outcome measure (math test performance) and the potential mediators (emotionality and worry). Gender was significantly related to math test performance in all grades, replicating the above ANOVA findings. In regard to the relation between gender and emotionality and worry, results showed that at the fourth
grade level, gender was not related to either of these anxiety components (emotionality: $b = -.32, s_b = .60, p > .05$; worry: $b = -.15, s_b = .08, p = .07$). However, at the eighth grade (emotionality: $b = -.29, s_b = .06, p < .01$; worry: $b = -1.88, s_b = .45, p < .01$) and twelfth grade (emotionality: $b = -.16, s_b = .05, p < .01$; worry: $b = -1.34, s_b = .35, p < .01$) gender was significantly related to both of these variables. Therefore, mediation analyses were continued at both the eighth and twelfth grades, but not at the fourth grade level.

First I present the results for the eighth grade students. Following the analyses of preliminary requirements for mediation that was run for these students, a regression model that included gender, emotionality, and worry as predictors of proportion correct on the math test was run. The findings revealed that gender was no longer a significant predictor of math test performance when the two anxiety measures were included in the regression model ($b = .03, s_b = .04, p = .52$) (Figure 6). To complete the mediation analysis, I used bias-corrected bootstrapping to estimate confidence intervals for the mediated effect – both for the full analysis and each of the mediators individually. The total indirect effect was significant with a point estimate of .08 and 95% bias-corrected confidence intervals of .04 to .13. Since the confidence interval does not overlap with 0, this was a significant indirect effect. In addition, the product of the path coefficients ($a_{wor} \times b_{wor}$) for the indirect path from gender to math test performance through worry was significant with a point estimate of .05 and a 95% bias-corrected confidence interval of .01 to .10. The test of the path coefficients for the indirect relation between gender and math test performance through emotionality ($a_{emo} \times b_{emo}$), however, was not significant with a point estimate of .03 and a 95% bias-corrected confidence interval of -.004 to .09.
Thus, it appears that worry, and not emotionality, significantly contributes to the indirect relation between gender and math performance.

Figure 6

Results of the Mediation Analysis, Eighth Grade

In the twelfth grade, findings were generally parallel to those at the eighth grade level (see Figure 7). The only difference was that gender was still a significant predictor of math performance when the anxiety measures were included in the regression model, though the relationship was weakened ($b = .08$, $s_b = .04$, $p = .03$). However, the total indirect effect was still significant with a point estimate of .04 and 95% bias-corrected confidence intervals of .01 to .08. In addition, the indirect relation from gender and math test performance through worry was significant with a point estimate of .03 and a 95% bias-corrected confidence interval of .01 to .07. Emotionality was not significant with a
point estimate of .01 and a 95% bias-corrected confidence interval of -.01 to .03. Again, it appears that worry, and not emotionality, significantly contributes to the indirect relation between gender and math performance.

Figure 7

*Results of the Mediation Analysis, Twelfth Grade*

![Diagram](image)

---

**Anxiety as a mediator of stereotype threat effects on math performance.** The fourth research question was about how stereotype threat and state anxiety were together related to gender differences in math performance. A mediation model was proposed, such that potential stereotype threat effects on girls’ math performance would be mediated by anxiety (worry and emotionality). Results reported above indicate that there were no such stereotype threat effects at any grade level, therefore it was not possible to test anxiety as a mediator of the impact of stereotype threat on math performance.
However, I was able to assess whether stereotype threat led to higher anxiety for girls, even though this higher anxiety could not have led to poorer math performance (since girls in the stereotype threat condition did not perform more poorly than girls in the no-threat condition). Regression results showed that girls in the stereotype threat condition were not more anxious before the test than girls in the no-threat condition at any grade level. This was the case for both the worry and emotionality components of state test anxiety (worry: fourth grade $b = 0.12, s_b = 0.12$; eighth grade $b = 0.07, s_b = 0.55$; twelfth grade $b = -0.32, s_b = 0.48$); emotionality: fourth grade $b = 0.14, s_b = 0.12$; eighth grade $b = -0.12, s_b = 0.08$; twelfth grade $b = 0.02, s_b = 0.07$).

**Discussion**

Study 1 examined gender differences in math performance, stereotype threat effects, and the role of anxiety in math performance in children of different ages, using comparable measures across age groups. The results of this study provided information relevant to the four research questions outlined above. The first research question concerned gender differences in math performance. Unlike the results from some prior research, there were gender differences in math performance at all grade levels and not only later in the school years (Hyde, Fennema, & Lamon, 1990; Lindberg et al., 2010). This inconsistency with the broad gender differences literature can potentially be explained by the fact that this assessment was quite difficult and the students were of high ability. When considering the literature on this type of assessment with high performing students, gender differences seem to be more robust. Some researchers (e.g.,
Gibbs, 2010) argued that gender differences in complex math problem solving may appear at earlier grade levels than on general assessments.

The second research question concerned the existence of stereotype threat, and the findings provided no evidence of a stereotype threat effect at the fourth, eighth, and twelfth grades. There are a number of possible reasons why stereotype threat effects did not occur. First, it is possible that the activation of stereotype threat used in this study was not strong enough to induce stereotype threat effects. The finding that stereotype threat also did not lead to heightened anxiety in girls lends support to this possibility. However, it should be noted that stereotype threat activation methods used in some other studies appeared to be even more subtle and nevertheless produce stereotype threat effects. It is also possible that stereotype threat effects, at least in the case of subtle manipulations, do not manifest themselves until later in development – an issue addressed further in Study 2. It should be noted that prior literature has produced mixed results with regard to stereotype threat effects in children and adolescents. Since there were no stereotype threat effects, I was not able to investigate the potential role of anxiety as a mediator.

I was able to address the third research question concerning the potential role of anxiety as a mediator of the relation between gender and math performance. The findings at the eighth and twelfth grade levels showed that the emotionality component of state anxiety, despite a significant zero-order correlation with both gender and math performance, did not act as a mediator when entered into the mediation model with worry. However, the worry component acted as a mediator for the relation between gender and math performance at both grade levels. These results fit with the general
pattern found in the literature indicating that there are gender differences in both components of state anxiety and that there is a relation between these components and performance (Hong & Karstensson, 2002; Miller & Bichsel, 2004). Furthermore, some researchers suggested that worry is more related to math performance than emotionality, which is supported by the present findings (Deffenbacher, 1980; Hembree, 1988; Kim & Rocklin, 1994).

These results differ, however, from the findings of other mediational studies. One study found no evidence for mediation with a trait math anxiety measure that combined emotionality and worry (Casey et al., 1997). Another study found that emotionality (measured after the test) was a significant mediator, though worry was not measured in this study (Osborne, 2001). The differing results are likely due to the fact that the current study measured state anxiety immediately prior to testing, which may be very different from both trait anxiety and state anxiety as measured after testing, which may be more of a reflection on performance. In addition, the current study is the first one examining both components (worry and emotionality) by including them separately in the mediation model. Thus, the novel finding of this study is that the relation between gender and math performance in eighth and twelfth graders is mediated by state worry immediately prior to math testing, the aspect of anxiety that is related to the cognitive processing of anxious thoughts. It appears that the physiological reaction to an evaluative situation is less important to performance than the worry-related thoughts and concerns that arise during testing. This lends support to the possibility that these thoughts tax working memory resources thus leading to poorer performance (Eysenck & Calvo, 1992).
The findings for fourth graders were different from both eighth and twelfth graders. There were gender differences in performance, but no gender differences in either component of anxiety. The lack of a gender difference in emotionality may be due to the fact that the measure was not very reliable for this age group ($\alpha = .530$). This low reliability can be interpreted as a poor measure of anxiety for fourth graders, most likely due to the higher vocabulary needed to understand all of the items. However, the fact that fourth graders’ responses to the worry measure are reliable suggests that anxiety is a construct that is measureable at the fourth grade level, but that the emotionality component may require more age appropriate vocabulary terms related to anxiety, or an entirely different way of measuring (e.g., nonverbal indicators of emotionality).

Given the developmental trends observed in Study 1, in the next study I further pursued this investigation of the nature of gender differences in an older sample. I investigated gender differences in college students and also decided to again manipulate stereotype threat, because there has been much research showing consistent stereotype threat effects at this age (see Nguyen & Ryan, 2008). In addition, Study 2 was designed to further expand the investigation of the nature of gender differences in math performance. Study 1 established mediation by an affective factor, worry, but still left open the question of the underlying cognitive mechanism for worry’s relation to math performance. In the next study, I pursued working memory as possible mediator of the effect of worry on math performance. Also, if stereotype threat effects are indeed found in this population, I will test a meditational model for stereotype threat effects on math performance that includes anxiety and working memory performance. Since the college
population tends to be impacted by stereotype threat, I will be able to more easily examine the mechanisms through which stereotype threat impacts performance.
Chapter 4: Study 2

Study 2 aimed to further investigate the nature of gender differences in math by examining both affective and cognitive factors related to math performance. Much literature has suggested that anxiety influences performance by interfering with working memory (Eysenck & Calvo, 1992; Owens et al., 2008). That is, when anxiety is increased, an individual monitors this anxiety, which takes away working memory resources, making it more difficult to use working memory resources for the mathematics task. Combined with Study 1 findings indicating that worry mediates the relation between gender and math performance, this explanation suggests that girls’ greater levels of worry may take up working memory resources which in turn, may lead to gender differences in math performance. This model has never been evaluated empirically. Therefore, I specifically tested whether worry and working memory act as a mediating chain from gender to math performance (Figure 1, p. 12).

Since there were no stereotype threat effects found in Study 1, which made it impossible to test for mediators of this effect, in Study 2, stereotype threat effects are examined in college students, who are more likely to be impacted by stereotype threat effects (Nyugen & Ryan, 2008; Spencer et al., 1999). If stereotype threat effects are indeed found, I will test a model looking at whether anxiety and working memory performance create a mediating chain of relations from stereotype threat to math performance (Figure 2, p. 33).

Study 2 employed a 2 (condition: stereotype threat, no-threat) x 2 (gender: male students, female students) experimental design. Stereotype threat was manipulated
between subjects. In addition, Study 2 examined the mediation model with anxiety and working memory for gender differences as well as for stereotype threat effects. These models were tested through statistical mediation analyses, which involve a series of regression analyses (Dearing & Hamilton, 2006; Hayes, et al., 2011). The key research questions examined in Study 2 were as follows:

1) Are there gender differences in college students’ math test performance?
It is expected that male college students will have higher performance on the math test than female students. This hypothesis is based on research that suggests that gender differences in math in general usually emerge in high school and continue to exist in college (Hyde, Fennema, & Lamon, 1990; Lindberg et al., 2010).

2) Does stereotype threat with regard to females and math impact the math test performance of college students?
It is predicted that women in the stereotype threat condition will perform more poorly compared to women in the no-threat condition and that men will not differ in their performance across conditions. This is based on the large body of literature showing stereotype threat effects in college women (Nguyen & Ryan, 2008; Spencer, et al., 1999).

3) Can gender differences in math performance be explained by women’s heightened worry, which decreases their working memory capacity? In other words, do worry and working memory create a mediating chain between gender and math performance?
Study 2 adds to the mediation model that was found to be significant in Study 1 by further testing whether working memory mediates the relation between worry and math
test performance. Based on the findings of past research concerning bivariate relations involved in this mediation model, it is proposed that gender differences in math will be partially explained by girls’ heightened worry and the subsequent deficit in working memory resources (Hong & Karstensson, 2002; Eysenck & Calvo, 1992; Owens et al., 2008; Raghubar et al., 2010). Worry is measured in two ways – with a thought listing measure and worry scale. Two types of working memory (verbal and visual) are measured due to inconsistent findings with regard to the component related to anxiety and math performance (Gathercole et al., 2004; Holmes et al., 2008). Only potential mediators that are correlated with both gender and math performance are tested in the model.

4) If there are stereotype threat effects on females’ math performance, can this effect be explained by females’ increased worry, which interferes with working memory? In other words, do worry and working memory create a mediating chain between stereotype threat and math performance?

It is expected that stereotype threat effects will be mediated by worry and then working memory such that stereotype threat leads to heightened worry in female students, which leads to a deficit in working memory resources which leads to poorer math performance. This analysis only includes female students because stereotype threat is not expected to impact the math performance of males. Thus females in the stereotype threat condition are compared to females in the no-threat condition. This hypothesis is based on research that shows that stereotype threat is related to worry and working memory (e.g., Cadinu et al., 2005; Schmader & Johns, 2003). In addition, research shows that worry is related to
working memory performance and that working memory performance is related to math performance (e.g., Crowe et al., 2007; Owens et al., 2008; Raghubar et al., 2010). Some work has even found that worry or working memory alone acted as a mediator of stereotype threat effects, but few studies have looked at these mediators together (Brodish & Devine, 2009; Cadinu et al., 2005; Johns et al., 2008; Schmader & Johns, 2003).

Method

Participants. Participants were 90 undergraduate students from a competitive Northeastern university. The sample included 65 women and 25 men (the unequal sample size is addressed and taken into account later when discussing the analyses). On average students were aged 19 years 11 months. Sixty-seven percent of students took Calculus in high school and 79% had taken at least one math class during college. Students were recruited by contacting college professors and consent was obtained from students prior to testing.

Materials. Participants completed the two measures of worry, the thought listing measure and the worry scale (Appendix D), as well as the twelfth grade mathematics test used in Study 1 (Appendix H). The emotion list was not used in this study because it was not a significant mediator of gender differences in math at any grade level in Study 1. Although in Study 1 the thought listing measure showed a floor effect, it was administered in Study 2 because this measure was previously used successfully with college students who may be more likely to report relevant information during a thought listing task (Cadinu et al., 2005). Two working memory measures were also administered to students.
Stereotype threat manipulation. The stereotype threat manipulation was similar to that in Study 1, but adapted for college students. The answer choices were the same as those for twelfth grade. The students in the stereotype threat condition read:

At one university, the male students were much better at math than the female students. The professors at the university chose the 30 students with the highest math test scores to represent the school in a national math competition. Twenty-seven of the students were male and three were female. What is the ratio of the number of males to the number of females among those who were chosen for the math competition?

The students in the no-threat condition read:

At one university, students were invited to participate in a service trip abroad, but there were only 30 spots available. The professors running the trip chose 27 students from one program and three students from another program. What is the ratio of the number of students from the first program to the number of students from the second program among those going on the service trip?

Working memory. Two dual-task working memory measures were administered to the students on a computer: a verbal working memory measure (a word recall task) and a visual working memory measure (a spatial recall task). Both of these measures were used because there is evidence suggesting that both of these components may be related to anxiety about math performance (Beilock et al., 2007; Crowe et al., 2007; Hadwin et al., 2005; Holmes & Adams, 2006; Shackman et al., 2006).
Verbal working memory. This measure involved word recall and was based on the listening recall assessment from the Automated Working Memory Assessment (Alloway, Gathercole, & Pickering, 2004). The student heard a sentence and judged whether it was true or false. After they heard 1 – 6 sentences in a row that they judged true or false, they were asked to state the last word of each of the sentences that they heard in order. For example, the student heard “Magazines have pages” then said true or false, then heard “Apples play football” and said true or false, and then they were asked to say the last word of each of the sentences in order “pages, football.” Students had to recall the words in the appropriate order to be considered correct. The researcher used the computer keyboard to indicate both whether the student said true or false and the accuracy of recall of the last words in the sentences in order. The task was divided into six blocks: the first block consisted of trials which included only one sentence, the second had trials with two sentences, etc. up to six sentences. Each block had a maximum of six trials, but the number of trials a student completed within a block varied. Due to the ease of trials with 1, 2, or 3 sentences, each student only did one trial from each of these first three blocks. For the remaining blocks (4, 5, or 6 sentences) the students completed trials within the block until they either (1) got 4 trials correct or (2) got 3 trials incorrect. If the student got 4 trials correct, they moved onto the next trial block without completing the last 2 trials in that block. However, if the student got 3 or more trials incorrect in the block, the task ended.

Students received two scores on the task: a processing score (the number of accurate true/false judgments) and a recall score (the number of times they correctly
recalled the words in order). Analyses were conducted with the recall score because it measures memory capacity while processing other information.

*Visual working memory.* This measure was based on the spatial recall task, also from the Automated Working Memory Assessment (Alloway, et al., 2004). This task followed a parallel structure to the word recall task, but with visual stimuli instead of verbal stimuli. Two shapes were presented next to each other. The participant had to determine if the shape on the right was the same or opposite of the shape on the left. The shape on the right was rotated either 0, 120, or 240 degrees and was presented with a red dot at the top of the shape. After participants saw 1 – 7 pairs of shapes in a row, they had to point to one of three dots (at 0, 120, and 240 degrees) that matched the dot location at the top of the shapes in the order in which they saw them. The administration procedure for the trial blocks was parallel to that in the verbal working memory task. Again participants’ recall scores, which were based on their correct recall of the locations of the dots, were used in analyses.

**Procedure.** Students were randomly assigned to either the stereotype threat condition or the no-threat condition and tested individually in an office at their university. Instructions given to the students were parallel to those in Study 1 except that participants were told that they would be taking two memory assessments as well.

The procedure for Study 2 is outlined in Figure 9. Step 1 involved the sample word problem that activated stereotype threat or not. In Step 2 students completed two anxiety measures (thought listing measure and worry scale). During Step 3 students completed the verbal and visual working memory measures (with the order of the
measures counterbalanced). In Step 4, students were reminded of the sample math problem they had done, to reinforce the stereotype threat manipulation for those in the stereotype threat condition. At this point, as in Study 1, the students in the no-threat condition were told that they were going to do some math problems, whereas students in the stereotype threat condition were told they would be taking a math test. In Step 5 students completed the math test.

Figure 8

Study 2 Procedure

Results

Preliminary analyses. In the preliminary analyses, the data were inspected to examine normality, test for outliers, and measure reliability.
**Normality check.** The examination of the shape of the distribution for the variables showed that the distributions for all scales were relatively normal.

**Outliers.** Outliers were identified using the method described in Study 1, that is a combination of the interquartile range method and the residual statistics for the regression analyses. Two cases (1 male, 1 female) fell outside of these cutoffs and were not included in statistical analyses. Thus, analyses were run on the remaining sample of 88 students.

**Reliability.** For the thought listing measure, interrater reliability was calculated based on the coding of two independent raters. Agreement occurred 95% of the time. For the worry scale and the math test, Cronbach’s alphas were computed as a measure of internal consistency. The reliability for the worry scale was high ($\alpha = .780$). The reliability for the math test was moderate ($\alpha = .614$), and slightly lower than the reliability for the test in the 12th grade ($\alpha = .639$). Again, since the test was specifically designed to measure a broad range of concepts with a small number of items, a high reliability would not be expected.

**Descriptive statistics.**

**Means and standard deviations.** Table 9 shows descriptive statistics for the relevant measures.
Table 9

*Descriptive Statistics (Means and Standard Deviations) by Gender and Stereotype Threat Condition*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Female Students</th>
<th>Male Students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stereotype Threat</td>
<td>No-Threat</td>
</tr>
<tr>
<td>Worry Scale</td>
<td>9.74 (1.97)</td>
<td>10.52 (2.18)</td>
</tr>
<tr>
<td>Worry - Thought Listing</td>
<td>0.68 (0.73)</td>
<td>0.89 (0.82)</td>
</tr>
<tr>
<td>Verbal Working Memory</td>
<td>22.97 (4.42)</td>
<td>22.24 (3.88)</td>
</tr>
<tr>
<td>Visual Working Memory</td>
<td>26.09 (6.19)</td>
<td>25.52 (6.71)</td>
</tr>
<tr>
<td>Math Score</td>
<td>0.38 (0.21)</td>
<td>0.37 (0.14)</td>
</tr>
</tbody>
</table>

Note. Scale for the worry scale is from 4 to 16; Scale for thought listing measure is frequency; Scale for verbal working memory is from 0 to 36; Scale for visual working memory is from 0 to 42; Scale for math score is proportion correct.

**Correlations.** Correlations between gender, scores on the worry scale, thought listing scores, verbal working memory, visual working memory and math test scores are presented in Table 10. Most variables were correlated with one another, however there are a few exceptions. The thought listing measure of worry was not correlated with either working memory measure. In addition, verbal working memory was also not correlated with gender or math test scores.
Table 10

*Correlations among Measures*

<table>
<thead>
<tr>
<th></th>
<th>Gender</th>
<th>Worry Scale</th>
<th>Thought Listing</th>
<th>Verbal WM</th>
<th>Visual WM</th>
<th>Math Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worry Scale</td>
<td>-.481**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thought Listing</td>
<td>-.220*</td>
<td>.501**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal WM</td>
<td>.063</td>
<td>-.247*</td>
<td>-.110</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual WM</td>
<td>.457**</td>
<td>-.412**</td>
<td>-.163</td>
<td>.326**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Math Score</td>
<td>.404**</td>
<td>-.475**</td>
<td>-.358**</td>
<td>.192</td>
<td>.477**</td>
<td>1</td>
</tr>
</tbody>
</table>

**Main statistical analyses.**

*Gender differences and stereotype threat effects.* In order to investigate gender differences and the effect of stereotype threat on math performance, a Factorial ANOVA was run. This analysis addressed the first two research questions. There were two between subject variables: condition (stereotype threat, no-threat) and gender (male, female). The dependent variable was the proportion of items answered correctly on the math test. Similar to Study 1, results were parallel when using both the proportion correct out of all twelve items and the proportion correct out of items attempted, and only the results of the former analyses are presented.

It was predicted that there would be a gender difference in math performance regardless of condition, which would be indicated by a significant main effect of gender.
In addition, it was predicted that stereotype threat would impact the math performance of female, but not male, students, which would be indicated by a significant interaction between gender and condition.

Similar to Study 1, unequal sample sizes had the potential of leading to problems in running an ANOVA due to potential effects of heterogeneity of variance. Again, the Levene’s test showed null results, thus the unequal sample sizes did not present a problem with the subsequent analysis (Keppel & Wickens, 2004). The results revealed a main effect of gender \(F(1, 83) = 16.28\ p < .001\), but not condition. The interaction between gender and condition was not significant. The main effect of gender showed that male students performed better than female students (Male students \(M = 0.56\), Female students \(M = 0.37\); \(d = 0.95\)). Thus, there were significant gender differences in math performance but no evidence of stereotype threat effects because there was no significant two-way interaction between stereotype threat condition and gender.

\textit{Worry and working memory as mediators of gender differences in math performance.} Since it has been established that there is a gender difference in math performance, the next step was to address the third research question by testing whether worry and working memory act as a path through which gender differences in math performance occur. In statistical terms, this pattern is captured in the model in which worry and working memory mediate the relation between gender and math performance (Hayes, et al., 2011). Figure 10 illustrates this mediation model. The critical path for the present analysis is the path from gender to worry \((a_1)\), worry to working memory \((a_3)\), working memory to math performance \((b_2)\). Note that, as in Study 2, stereotype threat
condition was included as a covariate in order to remove any variance accounted for by the differences in experimental condition.

Recall that students completed two measures of working memory and two measures of worry and that any combination of these could be used in testing the mediation model. However, based on the correlations reported above, only the worry scale and the visual working memory score were used in the model. The thought listing measure of worry was excluded because it did not have a significant correlation with either measure of working memory. In addition, verbal working memory could not be tested in the model because it was not significantly correlated with gender or math performance. Thus, these measures clearly would not have acted as significant parts of the proposed mediation model.
Figure 9

*Proposed Mediation Model for the Relation between Gender and Math Test Performance, College Students*

Analytic strategy. To test for mediation, a series of regression models were run. The first two regression analyses tested whether the predictor (gender) was related to the outcome measure (math performance) and the first potential mediator (worry). Next, a regression model that included gender and worry as predictors of visual working memory performance was run. The last regression analysis tested whether math performance was predicted from gender, worry, and visual working memory. Figure 10 depicts the model and the coefficients that represent each effect from these three analyses. Mediation is likely evident if the relation between gender and math test performance (c) disappears or significantly decreases after the addition of worry and visual working memory into the
model \((c')\). In other words, if there is a difference between the total effect of gender and the direct effect of gender (once worry and visual working memory are taken into account), this indicates that the indirect effect was likely significant.

To complete the mediation analysis, I again used bias-corrected bootstrapping to estimate confidence intervals for the mediated effect (Dearing & Hamilton, 2006; Preacher & Hayes, 2008; Hayes, et al., 2011). The analyses were run with the MED3C SPSS macro created by Preacher and Hayes. This method was recommended by Taylor, MacKinnon and Tein (2008) who conducted a Monte Carlo study comparing the Type I error, power, and coverage of six different methods for estimating this type of mediation model (a three-path mediation model). Using this method, the total indirect effect as well as three specific indirect effects were examined. These indirect effects were tested by calculating the product of the path coefficients between the predictor and mediator(s) and between the mediator(s) and the outcome. The three specific indirect effects are (1) the indirect effect of gender on math performance through worry \((a_1b_1)\) (2) the indirect effect of gender on math performance through visual working memory \((a_2b_2)\), and (3) the indirect effect of gender on math performance through worry and then visual working memory \((a_1a_3b_2)\). The test of the total indirect effect is the sum of the three specific indirect effects \((a_1b_1 + a_2b_2 + a_1a_3b_2)\) and is equivalent to testing the difference in the total effect of the predictor on the outcome \((c)\) and the direct effect of the predictor on the outcome \((c')\). As in Study 1, significance was assessed by examining the 95\% confidence intervals for each indirect effect to see if they overlap with 0 (Preacher & Hayes, 2004).
Estimation of the mediation model. Results of this mediation analysis are visually presented in Figure 11. To begin, I first ran two regression analyses to test whether gender was related to both the outcome measure (math test performance) and the potential first mediator of the model (worry). Replicating the findings of the above ANOVA, gender was significantly related to math test performance \((b = 0.19, s_b = 0.05, p < .01)\). Further, regression results showed that gender was significantly related to worry \((b = -2.48, s_b = 0.49, p < .01)\).

The next regression analysis for this model included gender and worry as predictors of visual working memory performance. The findings revealed that both gender \((b = 5.12, s_b = 1.65, p < .01)\) and worry \((b = -0.75, s_b = 0.32, p < .05)\) were significant predictors of visual working memory. The final regression analysis run was a regression predicting math performance from gender, worry, and visual working memory performance. The results showed that gender was no longer a significant predictor of math performance \((b = 0.06, s_b = 0.05, p > .05)\) but both worry \((b = -0.03, s_b = 0.01, p < .01)\) and visual working memory \((b = 0.01, s_b = 0.003, p < .01)\) were significant predictors of math performance.

To complete the mediation analysis, I used bias-corrected bootstrapping to estimate confidence intervals for the mediated effect – both for the full analysis and each of the three specific indirect effects (through worry only, visual working memory only, and through both worry and visual working memory). The total indirect effect was significant with a point estimate of .13 and 95\% bias-corrected confidence intervals of .07 to .19. Since the confidence interval did not overlap with 0, this was a significant
indirect effect. In addition, the product of the path coefficients \(a_1b_1\) for the indirect path from gender and math test performance through worry was significant with a point estimate of .07 and a 95% bias-corrected confidence interval of .01 to .12. The test of the path coefficients for the indirect relation between gender and math test performance through visual working memory \(a_2b_2\) was significant with a point estimate of .05 and a 95% bias-corrected confidence interval of .01 to .10. Most importantly, the indirect effect of gender on math performance through both worry and visual working memory \(a_1a_3b_2\) was significant with a point estimate of .02 and a 95% bias-corrected confidence interval of .001 to .04. Thus, there was evidence of a mediating chain from gender to math performance through worry and then visual working memory.

Figure 10

*Results of the Mediation Analysis, College Students*
Anxiety and working memory as mediators of stereotype threat effects on math performance. As in Study 1, results indicated that there were no significant stereotype threat effects, therefore research question 4, in regard to worry and working memory as mediators of the impact of stereotype threat on math performance, could not be tested. However, I did assess whether stereotype threat led to higher worry or poorer working memory performance for girls. Regression results showed that female students in the stereotype threat condition were not higher in worry as measured by both the thought listing measure ($b = -0.22, s_{b} = 0.20, p > .05$) and the worry scale ($b = -0.78, s_{b} = 0.53, p > .05$) than girls in the no-threat condition. Similar results were obtained for the relation between stereotype threat and both verbal ($b = 0.73, s_{b} = 1.06, p > .05$) and visual working memory performance ($b = 0.57, s_{b} = 1.63, p > .05$).

Discussion

In this sample of college students, there were robust gender differences on the math test, with male students outperforming female students. This finding fits with much research showing a gender difference in math during the college years (Hyde, Fennema, & Lamon, 1990; Lindberg et al., 2010). At the same time, the finding of no stereotype threat effects is inconsistent with much of the prior research on college students (Nguyen & Ryan, 2008; Spencer et al., 1999). It is possible that the stereotype threat manipulation that was used in this study was not strong enough to produce stereotype threat effects, even in a college sample. Another potential reason is the selection of the sample. It is possible that many of the students were not highly identified with math and therefore were unlikely to be affected by stereotype threat (Keller, 2007; Steele, 1997). Most of the
students had little college math course experience and were majoring in social science fields.

Since there was no evidence of stereotype threat, the mediators of stereotype threat could not be tested. However results showed that there were no stereotype threat effects on either measure of worry or either type of working memory. These results show that the stereotype threat effect was not only unrelated to math performance, but that it also did not have an impact on other related variables that have been found to be affected by stereotype threat in the literature (Schmader & Johns, 2003; Cadinu et al., 2005).

The inclusion of both types of working memory measures (verbal and visual) allowed me to identify the type of working memory that was most strongly related to gender differences in math performance. The results showed that verbal working memory did not act as a part of the mediation model. More specifically, verbal working memory was related to worry (just as visual working memory); however, in contrast to visual working memory, verbal working memory was not related to math performance. In the context of the processing efficiency theory, one would predict that that worry would be correlated with both verbal and visual working memory, and in fact much research supports this assertion (Crowe et al., 2007; Hadwin et al., 2005). However, with respect to math performance, the findings differ in implicating either verbal or visual working memory or both. Given the nature of the math problems included in the present research, it appears that visual working memory skills are more heavily tapped when students solve challenging math problems, a finding similar to that reported by Holmes and Adams (2006).
The finding that the worry component of anxiety and visual working memory created a mediating pathway from gender to math performance is the key novel finding of the present study. It suggests that gender differences in performance occur because of increased worry in female students, which taxes their visual working memory, which, in turn, leads them to perform more poorly on the math assessment. This result fits with prior work that showed evidence for the individual pathways involved in the model (e.g., the relation between gender and anxiety, anxiety and working memory, or working memory and math performance). The results do somewhat differ from those of other meditational studies that do not test the entire mediation model (Casey et al., 1997; Osborne, 2001; Owens et al., 2008) – this issue is addressed in the general discussion. However, this is the first empirical test of the entire model that simultaneously tested for the meditational chain from gender to worry to working memory to math performance.
Chapter 5: General Discussion

The main goal of this dissertation research was to examine the cognitive and affective factors underlying gender differences in math performance. This investigation grew from research on gender socialization (Eccles & Jacobs, 1986). As girls are socialized, they learn about the stereotype that girls are not good at math, and this exposure may lead them to be anxious about their math performance and to be susceptible to stereotype threat effects. Existing research suggests that heightened anxiety may have negative consequences for cognitive processing, which in turn, would lead to poor performance on math tests. Both lower levels of performance and the anxiety itself may deter female students from choosing educational paths leading to STEM careers. In order to address the shortage of female professionals in STEM fields, it is critical to better understand whether the affective and cognitive factors suggested in the literature as potentially underlying gender differences in math achievement indeed form a mediating chain from gender to math performance.

The current study addressed a number of questions that remain open based on the existing literature. First of all, although there is a general agreement that there is a substantial gender imbalance among adult professionals in STEM careers, there is currently much debate over whether gender differences still exist in math performance among school students. By administering difficult math assessments to high performing students, this study was able to examine gender differences in the areas where these differences are most likely to exist (Gibbs, 2010; Hyde, Fennema, & Lamon, 1990). Second, although there is much research suggesting that gender differences in math may
be explained by girls’ heightened anxiety which taxes their working memory, and leads to poor performance, there is a gap with respect to empirical investigations testing this potential mechanism (Hembree, 1988; Jarvis & Gathercole, 2003; Owens et al., 2008). The present research is the first to directly test this entire meditational relationship.

With regard to research on stereotype threat effects, there have been a number of studies with college students, but little work investigating this phenomenon in children and adolescents. The findings of the current literature with children and adolescents are mixed. In addition, the mechanisms underlying stereotype threat effects are not completely understood – both anxiety and working memory have been implicated as potential mediators. A number of studies have tested either anxiety or working memory as mediators of this relation, but they have rarely been tested in the same model (Beilock et al., 2007; Cadinu et al., 2005; Schmader & Johns, 2003). Thus, the present study broadened the research base by examining the occurrence of stereotype threat effects across development and testing potential mediators of this relation.

**Gender Differences in Math Performance**

Some researchers have recently argued that gender differences in math learning are virtually disappearing, and that small differences only appear in high school or in college (Hyde, Fennema, & Lamon, 1990; Hyde et al., 2008; Lindberg et al., 2010). Other investigators, however, argue that gender differences still exist as early as elementary school, but that they only appear on particular types of mathematical tasks (Gibbs, 2010). Specifically, math test items that are more complex and challenging tend to show small gender differences favoring males starting as early as first grade, whereas
some skills (e.g., computation) show a female advantage (Gibbs, 2010; Vasilyeva et al., 2009). Thus, it is possible that gender differences in math do not appear until later in development because earlier assessments include a larger portion of items that girls excel on whereas later assessments are more likely to assess the skills that show a male advantage. Some of the reported results showing no gender differences, such as Hyde and colleagues’ (2008) study on state-administered assessments, have been based on tests that generally do not include conceptually challenging items. Thus, the fact that the types of items showing a male advantage are not included in many current assessments may be the reason why gender differences are not found.

If there are indeed no gender differences in math performance, then the persisting underrepresentation of women in STEM careers may be largely due to cultural expectations and norms that lead women to make certain career choices (Eccles & Jacobs, 1986; Hyde, Fennema, & Lamon, 1990). On the other hand, if there are still differences in math skills at the earlier educational stages, which are not detected by current standardized assessments, it would highlight the importance of addressing affective and cognitive factors associated with math achievement early on. Thus, despite the large body of research on gender differences, it is particularly important to look specifically at the types of assessments that are more likely to reveal differences.

In order to contribute to this debate in the literature, an assessment that was designed to include the most challenging items from several national and international assessments was used in this study. The results showed that there were gender differences on the test across school age and college. These results highlight the fact that in more
complex math, even as early as fourth grade, boys outperform girls. By using aggregate math test scores, much research may gloss over the gender differences that exist in particularly difficult mathematics.

Interestingly, gender differences on the math assessments used in this research appeared to stay fairly consistent across elementary, middle and high school (0.5 < $d$’s < 0.6) but increase drastically in college ($d = 0.95$). This may be related to the fact that there are many fewer choices for math courses during compulsory schooling – all students are required to take at least one mathematics course every year. However, when students reach college they can choose to take as little as one math class or as many as 10-12 if they major in a math-intensive field. In the present sample of college students, even though there was no significant difference in the majors for males and females (most students majored in various social science disciplines, many from the school of education), there was a significant gender difference in the number of math courses that students had taken, though the means for both genders were quite low. It is possible that greater math anxiety in female school students leads them to be less likely to take math courses in college and this difference may be related to the larger gender effect found at this age (Eccles & Jacobs, 1986). In addition, it is possible that there is a cumulative effect of anxiety about math testing over time and this may lead to increasingly detrimental effect on performance as girls’ progress to more complex math.

**Lack of Stereotype Threat Effects**

Stereotype threat has been implicated in the literature as a potential explanation for gender differences in math performance. The present work adds to the literature by
examining stereotype threat effects across development using parallel experimental methods. The findings for school-aged students reported in previous studies were quite inconsistent, with some investigators finding stereotype threat effects, others finding them only at a particular grade level, and yet others not finding them at all (Ambady et al., 2001; Huguet & Regner, 2007, 2009; Muzzatti & Agnoli, 2007). With respect to college students, the findings tend to be more consistent in providing evidence of math stereotype threat effects (see Nguyen & Ryan, 2008). In the present research, there was no evidence of a stereotype threat effect across the age groups examined in the two studies: fourth-, eighth-, twelfth-graders, and college students.

In discussing potential reasons for the obtained results, it is important to consider the methodology of the present study in the context of other work on stereotype threat. Existing studies have utilized a variety of methods for activating stereotype threat. The methods range from subtle, implicit, manipulations aimed at activating students’ gender awareness (e.g., by having them mark their gender before taking the test) to more explicit, almost blatant, ways of activating the stereotype (e.g., by telling students prior to taking the test that girls are worse than boys in math and specifically on the kind of test that they are about to take). The particular method of activating stereotype threat used in the present study provided students with information consistent with the stereotype about girls and math.

In contrast to many other studies that activate the stereotype, the method utilized in the present research was less explicit. First, it was not specifically mentioned in the stereotype threat condition that there were gender differences in math in general; rather
children were provided with a situation that was consistent with the stereotype. Second, the activation was seemingly part of the normal math testing procedure. This type of activation method was chosen because it was more ecologically valid, since it is unlikely that students are ever directly told about gender differences in their normal environment. However, this activation method required students to make an inference about gender and math from their observation of the situation in the sample word problem – much like they do from observations in everyday life. In addition, this type of method has been used in prior research, which found stereotype threat effects at the eighth grade level, but not in younger students (Muzzatti & Agnoli, 2007). However, it is possible that this procedure was not strong enough to produce stereotype threat effects in this sample. Potentially this manipulation was too implicit. Many students may have read the sample word problem and not truly processed its contents, focusing more on the mathematics required to answer the math question. Thus the prompt may not have activated the stereotype threat about girls and math because students were ignoring the context of the word problems while solving it.

However, as indicated above, even though the stereotype threat manipulation may have been relatively weak in the present study, some other studies that have reported stereotype threat effects have also used very subtle manipulations. The discrepancies in findings concerning the effects of the math stereotype on female students’ performance may lead to questions about whether this particular stereotype threat has been decreasing over time. Yet, given the relative consistency of recent findings in older students, it is likely that this phenomenon still persists; however, it is not clear what conditions lead to
the activation of the stereotype threat. It would be beneficial for our understanding of this phenomenon to compare the various studies that have produced evidence of math stereotype threat to determine common types of prompts that lead to the activation of stereotype threat and that would be appropriate across many ages.

In addition to considering the features of experimental conditions that are likely to elicit stereotype threat, it is also important to consider the characteristics of participants that may be related to the likelihood of activating a particular stereotype. Much of the work that finds stereotype threat effects at the college level involves participants selected for high math identification (Schmader & Johns, 2003; Spencer et al., 1999) and some studies have found that being identified with math is important for stereotype threat effects to occur (Keller, 2007; Schmader, 2002). It is possible that the participants in the present study, including the college students, did not have high enough levels of math identification to be impacted by stereotype threat effects.

The majority of students in Study 2 were pursuing majors in the social sciences, suggesting that they may not be highly identified with math. With respect to Study 1, the participants were drawn from high-performing schools but the samples were not confined to students who had a particularly high identification with math. Interestingly, most of the work on math stereotype threat in children and adolescents has not selected for math identification and some have found stereotype threat effects at these earlier grade levels. Therefore, it is unclear whether the importance of math identification varies with age and if it may have affected the findings in regard to stereotype threat in the present study.
Worry as a Mediator of the Relation between Gender and Math Performance

Even though the present research did not reveal effects of the gender stereotype threat on math performance, it provided important information concerning the involvement of affective factors in the relation between gender and math. The findings of both studies indicated that the worry component of anxiety plays a role in gender differences in math performance. In particular, at the eighth and twelfth grades, as well as with college students, results showed that worry mediated gender differences in math performance. The bivariate relations involved in this model have been tested extensively in the literature, pointing to a relation between gender and anxiety as well as a relation between anxiety and math performance, with worry being more strongly related to performance than emotionality (Deffenbacher, 1980; Hembree, 1988; Hong & Karstensson, 2002; Kim & Rocklin, 1994; Miller & Bichsel, 2004). However, only two studies have assessed anxiety as a mediator of the relation between gender and math performance. One study measured state emotionality after the test and found that this did indeed mediate the relation between gender and math performance (Osborne, 2001). However, testing anxiety after the test is problematic as differences in the levels of anxiety at this point may reflect differences in students’ actual performance on the test. The second study, however, measured trait math anxiety combining emotionality and worry, and found null results (Casey et al., 1997). The research presented here differed from these studies by focusing on anxiety occurring immediately prior to the particular testing situation, and by separating out and testing both the worry and emotionality components of anxiety. Thus the overall pattern of relations is a novel finding but fits
with most past work in regard to the relation between gender, anxiety and math performance and the more prominent role of worry over emotionality.

The findings of Study 1 also revealed developmental differences in the measures of anxiety. Specifically, in the fourth grade there were no significant gender differences in either the emotionality or worry components of anxiety (though the effect for worry was marginally significant, \( p = .07 \)), despite a gender difference in performance. Past research on trait test anxiety has shown that the gender difference emerges in the lower elementary years (Hembree, 1988), but with respect to math anxiety, gender differences are usually not found until students are in middle school (Hyde, Fennema, Ryan, et al., 1990). The lack of a gender differences at the fourth grade level may be partly due to the fact that the available measure for emotionality may not be perfectly suited for this age group, due to lack of knowledge of some vocabulary words used, as evidenced by its low reliability (\( \alpha = .530 \)). In regard to worry, it is likely that the gender difference was not significant due to a lack of power (\( N = 68 \)), as the measure was quite reliable (\( \alpha = .885 \)) and there was a gender difference at the trend level.

Past research has found that the worry component of anxiety is more related to math performance than the emotionality component, however little work has simultaneously considered these two predictors at the same time and in the context of gender differences (Deffenbacher, 1980; Hembree, 1988; Kim & Rocklin, 1994). The results here show that both emotionality and worry were significantly related to gender and math performance when looking at zero-order correlations. However, when testing the mediation model, worry, but not emotionality, was a significant mediator of the
relation between gender and math performance at the eighth and twelfth grades. Once emotionality and worry had to compete for variance, only worry appeared to be the important mediator. Worry represents the cognitive component of anxiety that involves concerns about performing poorly. Based on these findings, it appears that when both emotional and cognitive components of anxiety are considered together, worry-related thoughts prior to testing appear to be more robust in explaining gender differences.

Further, in analyzing the nature of the affective processes that I was measuring, it is important to point out that our measure was explicitly meant to test state anxiety, (i.e., anxiety experienced during the particular testing situation), rather than trait math anxiety (i.e., a general tendency to feel anxious and less confident about math). Most researchers suggest that the two types of anxiety are related in that individuals with higher trait anxiety are more likely to have higher state anxiety. At the same time, the two types of anxiety are still viewed as two separate constructs, and there is work suggesting that state anxiety may have a stronger impact on performance than trait anxiety, potentially because of its proximity to the testing situation (e.g., O’Neil & Fukumura, 1992). This may explain discrepancies in some of the findings, in particular, the current finding that state anxiety mediates gender differences in math performance and earlier findings that trait math anxiety does not mediate these gender differences (Casey et al., 1997).

The findings of the present research add further support to the accumulating evidence implicating anxiety in poor math performance. The strongest tests of this relation can be provided by experimental studies examining whether changes in anxiety can lead to changes in math outcomes. There have been a few studies testing the causal
relation between test anxiety and test performance (Hembree, 1988; Wood, 2006).

Hembree (1988) found that anxiety during testing can be reduced through behavioral and cognitive-behavioral interventions and that this lower anxiety is followed by an improvement in test performance. It should be noted however, that this particular study involved a meta-analysis that examined both testing situations related to and unrelated to math. The effectiveness of these interventions combined with the findings of the current study suggests that interventions for anxiety may need to be targeted specifically for girls and specifically with regard to math. The use of these interventions can have potentially important consequences for girls’ achievement in mathematics.

**Worry and Visual Working Memory as Mediators of the Relation between Gender and Math Performance**

Finding that worry mediates the relation between gender and math performance is an important step towards a better understanding of the nature of gender difference in math. However, this finding raised the next question, concerned with explaining the role of anxiety in math performance. Several researchers have suggested that the relation between anxiety and test performance may reflect the negative effects of anxiety on working memory (Eysenck & Calvo, 1992; Owens et al., 2008). In the present work, I directly tested the path from anxiety to working memory to performance in explaining gender differences in math. This full set of relations has not been tested in prior work, but there is one study that explored the relation between anxiety, working memory, and math performance, although not in the context of gender differences. In this study, Owens and colleagues (2008) found that anxiety impacted math performance through its relation to
working memory. The results of the present research expand these findings by indicating not only that working memory plays a role in the relation between anxiety and math performance, but also that this set of relations mediates gender differences in math.

Existing research indicates that both verbal and visual working memory contribute to math performance (Holmes & Adams, 2006; Jarvis & Gathercole, 2003; Miller & Bichsel, 2004). However, in the present study, only visual working memory was related to math performance, which may have to do with the particular characteristics of the math test. Holmes and Adams (2006) found that the type of working memory related to math performance differed depending on the difficulty of the items. Specifically, visual working memory was related to performance on more difficult items, whereas verbal working memory was only related to performance on less difficult test items. It is possible that complex math problem solving relies more on abstract spatial reasoning. The particular math test used in the present study included a large number of items that required visualization and all the items were selected to represent high levels of difficulty, based on national and international assessments. Thus, it is not surprising that visual working memory was critical to performing well on this particular test. The difference between the present findings and the findings of Owens et al. (2008), which implicated verbal working memory, highlights the importance of considering the specific type of math problems when evaluating the role of both types of working memory in math performance.

The pattern of relations obtained in the present work suggests that interventions targeted at working memory may in fact improve math performance. A number of studies
have conducted experimental work with working memory interventions and found that working memory can be improved (e.g., Klingberg, Forssberg, & Westerberg, 2002; Thorell, Lindquist, Nutley, Gunilla, & Klingberg, 2009) and, furthermore, that there are subsequent gains in math performance (Holmes, Gathercole, & Dunning, 2009). These studies provide converging evidence linking the improvement of working memory skills to improved math performance.

Limitations and Future Directions

The null finding concerning stereotype threat effects in the present study leaves open several alternative interpretations. One possibility is that stereotype threat effects were in fact occurring in both the threat and no-threat conditions. Even though we did not observe a difference between these two conditions, there is still a possibility, pointed out by some researchers, that stereotype threat exists in any math testing situation (Smith & White, 2002; Steele, 1997). Perhaps the testing situation in general activated stereotype threat that occurs in everyday testing environments, regardless of the added manipulation. Thus, this may have led to the gender differences in math performance found at all ages, regardless of condition. In order to address this possibility, future research should employ not only stereotype threat and no-threat conditions, but also a stereotype nullification condition (Smith & White, 2002). This condition would include information that is inconsistent with the stereotype that males are better than females at math – either about girls and boys performing equally in math or about girls performing better than boys in math. If girls who are in a stereotype nullification condition do better than both girls in the stereotype threat and the no-threat condition, then this would provide evidence that
stereotype threat is occurring in normal testing situations and impacts performance in both the threat and no-threat conditions. This is indeed what Smith and White (2002) found with undergraduate students. Future research should investigate these three types of conditions with children and adolescents as well as college students.

In discussing the mediational pathway examined in present research, it must be acknowledged that both studies were correlational, and thus causal relations cannot definitively be established. Most likely, some relations here are bidirectional in nature (e.g., the relation between worry and math performance). Conceptually, mediation offers a potential mechanism underlying the relation between the predictor and outcome (Dearing & Hamilton, 2006). Establishing this type of relation is critical to the investigation of gender differences in math performance. The current research should be followed up by experimental investigations that test the model uncovered in the present study.

There were a number of methodological limitations to consider. First, results cannot be easily generalized to populations of students that differ from those in this study. The present findings were obtained with higher performing groups of students, and the results may be very different in other populations (Keller, 2007; Smith & White, 2001). Additionally, as discussed earlier, the stereotype threat manipulation utilized in this study may have been somewhat weak and results regarding stereotype threat effects could potentially be different with an alternative manipulation.

With regard to measuring anxiety, the thought listing worry measure administered in this study yielded little usable data at the fourth, eighth, and twelfth grades, thus it
appears that this measure is most appropriate for use in college samples. In addition, the measure of state emotionality was unreliable at the fourth grade level. This makes it difficult to make any strong conclusions about the relation between gender, anxiety and math test performance at this level. It is possible that anxiety does act as a mediator of gender differences in math at this age, but due to measurement issues, I was unable to capture this. Thus more work should be done to find suitable measures for anxiety at this age so as to better understand the processes involved in gender differences in math among elementary school students.

With regard to the math assessment, the test used in this study consisted of only twelve items (six within each content area) due to time constraints. The diversity of items selected within each content area (Number/Algebra and Geometry/Measurement) combined with the small number of items made it difficult to conduct analyses by content area. The present goal was to represent a broad range of skills, but it may be instructive to pursue this issue with a more extensive math test that would include a larger number of items for each separate content area in order to determine if gender differences or stereotype threat effects and the meditational models are different across content areas. For example, perhaps visual working memory is more important for Geometry/Measurement and less important for Number/Algebra.

**Implications**

The findings obtained in the present research show that gender differences on some mathematical tasks still exist, and that these specific gender differences develop at quite an early age. In addition, results suggest that it is through worry and visual working
memory that gender is related to math performance. Based on these findings, there are clear areas in which changes can be made to help girls perform better in mathematics, including interventions to relieve anxiety and improve visual working memory. These interventions will likely lead to subsequent math performance improvements. Anxiety, its impact on visual working memory, and the underperformance that it leads to, have substantial consequences for women’s involvement in STEM careers, which is problematic for the U.S. in this increasingly global economy.

In order to compete for STEM careers in the global market, the U.S. needs to optimize math and science learning in all students. There have been important shifts in both the national and international job markets with an increase of positions in STEM fields that will need to be filled over the next few decades. Thus, it is becoming more important for the U.S. to focus its resources on improving the education of students so that they may be successful in these careers. Until both men and women are encouraged to strive to reach higher achievements in mathematics, the U.S. will not be taking full advantage of its intellectual resources. Though women are attending college at higher rates than men, they are not entering and being retained in STEM majors at the same rate, and the differences become even more striking when considering advanced degrees (NSF, 2009). This is problematic, because although women can surely obtain a college degree and be successful, these patterns suggest that they may not be prepared for the shift in the job market toward more STEM careers. The findings of the present dissertation highlight the importance of alleviating anxiety and training visual working
memory skills as ways of improving girls’ math performance, which, in turn, may lead to higher levels of involvement in STEM careers.
References


Implications for mathematical development and mathematical curricula.  

*Educational Psychology, 26,* 339–366. doi:10.1080/01443410500341056


National Science Board (2010). Preparing the next generation of STEM innovators:


Appendices
Appendix A. Fourth Grade Stereotype Threat Manipulation Word Problems

Stereotype Threat Condition
At the Miller Elementary School, the boys were much better at math than the girls. The math teachers gave 10 prizes to the students with the best score on a math test. Nine of the prizes were given to boys and only one was given to a girl. What fraction of the prizes were given to boys?

No-Threat Condition
At the Miller Elementary School, students were invited to participate in a special field trip, but there were only 10 spots available. The teachers chose nine students from Ms. Fletcher’s class and only one student from Ms. Johnson’s class. What fraction of the students going on the field trip were from Ms. Fletcher’s class?

Answer Choices

A) \( \frac{1}{10} \)

B) \( \frac{9}{10} \)

C) \( \frac{1}{9} \)

D) \( \frac{9}{1} \)

E) \( \frac{10}{9} \)
Appendix B. Eighth Grade Stereotype Threat Manipulation Word Problems

Stereotype Threat Condition
At the Miller Middle School, the boys were much better at math than the girls. The math teachers chose the 20 students with the highest math test scores for the math team to represent the school at the statewide math competition. Eighteen of the students were boys and two were girls. What proportion of the students on the math team were boys?

No-Threat Condition
At the Miller Middle School, students were invited to participate in a special field trip, but there were only 20 spots available. The teachers chose 18 students from Ms. Fletcher’s homeroom and two other students from Ms. Johnson’s homeroom. What proportion of the students going on the field trip were from Ms. Fletcher’s homeroom?

Answer Choices

A) \( \frac{1}{10} \)

B) \( \frac{9}{10} \)

C) \( \frac{1}{9} \)

D) \( \frac{9}{1} \)

E) \( \frac{10}{9} \)
Appendix C. Twelfth Grade Stereotype Threat Manipulation Word Problems

Stereotype Threat Condition
At the Miller High School, the boys were much better at math than the girls. The teachers chose the 30 students with the highest math test scores for the math team to represent the school in a statewide math competition. Twenty-seven of the students were boys and three were girls. What is the ratio of the number of boys to the number of girls among those who were on the math team?

No-Threat Condition
At the Miller High School, students were invited to participate in a special field trip, but there were only 30 spots available. The teachers chose 27 students from Ms. Fletcher’s homeroom and three other students from Ms. Johnson’s homeroom. What is the ratio of the number of students in Ms. Fletcher’s homeroom to the number of students in Ms. Johnson’s homeroom among those going on the field trip?

Answer Choices

A) 1 to 10
B) 9 to 10
C) 1 to 9
D) 9 to 1
E) 10 to 9
Appendix D. Worry Scale

Please show how much you agree or disagree with each statement. The choices are from 1 “strongly disagree” to 4 “strongly agree.” Circle one of the four choices.

1) I feel very confident about my performance on the test I’m about to take.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>2</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>3</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>4</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>

2) I am worried that I may not do well on this test.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>2</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>3</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>4</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>

3) I feel that I may let myself down in my performance on this test.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>2</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>3</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>4</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>

4) I think that I may do quite well on this test.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>2</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>3</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>4</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>
Appendix E. Emotion List

Please circle the number that shows how much you are feeling each emotion right now.

<table>
<thead>
<tr>
<th>Emotion</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interested</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Afraid</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Frustrated</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Determined</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Ashamed</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Happy</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Upset</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Calm</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Unintelligent</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Excited</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Nervous</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Confident</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Enthusiastic</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Anxious</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Smart</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Uncertain</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
Appendix F. Fourth Grade Mathematics Test Items

1) What are all the whole numbers that make $4 \times \Box < 24$ true?
A) 0, 1, 2, 3, 4, 5, 6
B) 0, 1, 2, 3, 4, 5
C) 0, 1, 2, 3, 4
D) 0, 1, 2, 3
E) 6

2) The two number sentences shown below are true.

\[
\begin{align*}
\square - \bigcirc &= 6 \\
\bigcirc + \bigcirc &= 2
\end{align*}
\]
If both equations shown above are true, which of the following equations must also be true?

A) \(\square \times \bigcirc = \bigcirc\)
B) \(\bigcirc \times 2 = \bigcirc\)
C) \(\square + \square = 12\)
D) \(\square + \bigcirc = \square\)
E) \(\square - \square = \bigcirc\)
3) The objects on the scale above make it balance exactly. According to this scale, if △ balances ⬤⬤⬤, then □ balances which of the following?

A) ⬤
B) ⬤⬤
C) ⬤⬤⬤
D) ⬤⬤⬤⬤
E) ⬤⬤⬤⬤⬤

4) What do you have to do to each number in Column A to get the number next to it in Column B?

<table>
<thead>
<tr>
<th>Column A</th>
<th>Column B</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>5</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
</tr>
</tbody>
</table>

A) Subtract 12 from the number in Column A
B) Multiply the number in Column A by 5
C) Subtract 8 from the number in Column A
D) Add 8 to the number in Column A
E) Divide the number in Column A by 5
5) On a test, Hannah scored 8 points higher than Todd. On the same test, Hannah scored 7 points lower than Juanita.

\( H \) represents Hannah's score on the test.
\( T \) represents Todd's score on the test.
\( J \) represents Juanita's score on the test.

Based on the information above, which of the following must be true?

A) \( J < T \)
B) \( J < H \)
C) \( T > J \)
D) \( T < J \)
E) \( H > J \)

6) Which rule describes the pattern shown in the table?

<table>
<thead>
<tr>
<th>□</th>
<th>△</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
</tr>
</tbody>
</table>

A) □ + 5 = △
B) □ + □ = △
C) □ + □ + 1 = △
D) □ + □ + □ + 3 = △
E) □ + □ + 2 = △
7) This figure will be turned to a different position.

Which of these could be the figure after it is turned?

A)  

B)  

C)  

D)  

E)  

8) Based on the map above, about how many miles is the shortest route from Oakdale to Fenton?

A) 20  

B) 40  

C) 50  

D) 70  

E) 100
9) In this figure, how many small cubes were put together to form the large cube?

A) 6
B) 7
C) 8
D) 12
E) 24

10) Five children measured the width of a room by counting how many paces it took them to cross it. The chart shows their measurements.

<table>
<thead>
<tr>
<th>Name</th>
<th>Number of Paces</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miriam</td>
<td>11</td>
</tr>
<tr>
<td>Stephen</td>
<td>10</td>
</tr>
<tr>
<td>Erlane</td>
<td>8</td>
</tr>
<tr>
<td>Ana</td>
<td>9</td>
</tr>
<tr>
<td>Carlos</td>
<td>7</td>
</tr>
</tbody>
</table>

Who had the longest pace?

A) Miriam
B) Stephen
C) Erlane
D) Ana
E) Carlos
11) This picture shows a cube with one edge marked. How many edges does the cube have altogether?

A) 6  
B) 8  
C) 9  
D) 12  
E) 24

12) The grid below shows the locations of some places in Jim’s neighborhood.

Moving along the grid lines, what is the least number of blocks from Jim’s house to the school?

A) 9  
B) 7  
C) 6  
D) 4  
E) 3
Appendix G. Eighth Grade Mathematics Test Items

1) The table shows a pattern:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
</tr>
</tbody>
</table>

If the pattern shown in the table were continued, what number would appear in the box at the bottom of column B next to 14?

A) 19  
B) 23  
C) 27  
D) 29  
E) 31

2) If \( \frac{a}{b} = 70 \), then \( \frac{a}{2b} = \)

A) 35  
B) 68  
C) 70  
D) 72  
E) 140
3) Which of the following is the graph of the line with equation $y = -2x + 1$?

A) 

B) 

C) 

D) 

E) 

4) The length of a rectangle is 3 more than its width. If $L$ represents the length, what is an expression for the width?

A) $3 ÷ L$ 

B) $L - 3$ 

C) $L × 3$ 

D) $L + 3$ 

E) $L ÷ 3$
5) Which graph below represents the solution to the inequality below?
\[ 4x - 12 \geq x + 3 \]

A) 

B) 

C) 

D) 

E) 

6) The owner of a car dealership noticed a pattern in the weekly car sales, as shown in the table below.

| Weekly Car Sales |
|------------------|------------------|
| Week \((w)\)     | Number of Cars Sold \((s)\) |
| 1                | 12                |
| 2                | 18                |
| 3                | 24                |
| 4                | 30                |

For weeks 1 through 4, which of the following equations could represent the pattern of \(s\) cars sold during week \(w\)?

A) \(s = 6w\)  
B) \(s = 12w\)  
C) \(s = 6(w + 1)\)  
D) \(s = 6(w + 6)\)  
E) \(s = w + 11\)
7) In the figure above, a circle with center O and radius of length 3 is inscribed in a square. What is the approximate area of the shaded region?

A) 4  
B) 8  
C) 9  
D) 28  
E) 33

8) How many of the unit cubes above would it take to make the object below?

A) 15  
B) 16  
C) 30  
D) 32  
E) 45
9) The coordinate plane shown below has a figure in the third quadrant.

Which of the following shows the same figure after it has been reflected across the \( y \)-axis and then reflected across the \( x \)-axis?

A) 

B) 

C) 

D) 

E) 

10) If the string in the diagram is pulled straight, which of these is closest to its length?

A) 5 cm  

B) 6 cm  

C) 7 cm  

D) 8 cm  

E) 9 cm
11) In the triangle, what is the degree measure of $\angle ABC$?

A) 45
B) 110
C) 120
D) 135
E) 160

12) A square pyramid is shown below. What is the total number of edges in a square pyramid?

A) 4
B) 5
C) 6
D) 7
E) 8
Appendix H. Twelfth Grade and College Mathematics Test Items

1) Yvonne has studied the change in cost of tickets over time for her favorite sports team. She has created a model to predict the cost of a ticket in the future. Let $C$ represent the cost of a ticket in dollars and $y$ represent the number of years in the future. Her model is as follows.

$$C = 2.50y + 13$$

Based on this model, how much will the cost of a ticket increase in two years?

A) $5
B) $8
C) $13
D) $18
E) $26

2) The length of a rectangle is 3 more than its width. If $L$ represents the length, what is an expression for the width?

A) $3 \div L$
B) $L \div 3$
C) $L \times 3$
D) $L + 3$
E) $L - 3$
3) Which of the following is the graph of \(|2x - 5| \geq 3\) ?

A) 
\[ \begin{array}{c}
-4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 \\
\end{array} \]

B) 
\[ \begin{array}{c}
-4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 \\
\end{array} \]

C) 
\[ \begin{array}{c}
-4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 \\
\end{array} \]

D) 
\[ \begin{array}{c}
-4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 \\
\end{array} \]

E) 
\[ \begin{array}{c}
-4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 \\
\end{array} \]

4) For what value of \(x\) is \(8^{12} = 16^x\)?

A) 3
B) 4
C) 8
D) 9
E) 12
5) If \(3x + 2y = 11\) and \(2x + 3y = 17\), what is the average (arithmetic mean) of \(x\) and \(y\)?

A) 2.5  
B) 2.8  
C) 5.6  
D) 5.8  
E) 14

6) The table below shows a linear relationship between \(x\) and \(y\).

<table>
<thead>
<tr>
<th>(x)</th>
<th>(y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>−7</td>
<td>(a)</td>
</tr>
<tr>
<td>−3</td>
<td>10</td>
</tr>
<tr>
<td>−1</td>
<td>6</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>−6</td>
</tr>
</tbody>
</table>

What is the value of \(a\)?

A) −18  
B) −14  
C) 14  
D) 18  
E) 36
7) If triangles $ADE$ and $ABC$ shown in the figure above are similar, what is the value of $x$?

A) 4 \\
B) 5 \\
C) 6 \\
D) 8 \\
E) 10

8) A circle with diameter 10 centimeters is to be cut from a square of paper 10 centimeters on a side. Of the following, which is closest to the amount of paper left over after the circle is cut out?

A) 9 square centimeters \\
B) 21 square centimeters \\
C) 24 square centimeters \\
D) 69 square centimeters \\
E) 84 square centimeters
9) Semicircles are constructed on the sides of an equilateral triangle, as shown in the figure above. Of the following, which best approximates the sum of the lengths of the three darkened arcs?

A) 4.4258  
B) 4.7124  
C) 6.0000  
D) 6.7124  
E) 9.4258

10) In the xy-plane, a line parallel to the x-axis intersects the y-axis at the point (0, 4). This line also intersects a circle in two points. The circle has a radius of 5 and its center is at the origin. What are the coordinates of the two points of intersection?

A) (2, 1) and (2, -1)  
B) (3, 4) and (3, -4)  
C) (3, 4) and (-3, 4)  
D) (5, 4) and (-5, 4)  
E) (5, 0) and (-5, 0)
11) In the figure shown below, triangle $TUV$ is formed by joining the midpoints of the sides of equilateral triangle $QRS$. Triangle $WYZ$ is formed by joining the midpoints of the sides of triangle $TUV$.

If the area of triangle $QRS$ is 64 square inches, what is the area of triangle $WYZ$?

A) 1 square inch  
B) 4 square inches  
C) 8 square inches  
D) 16 square inches  
E) 64 square inches

12) How many of the unit cubes above would it take to make the object below?

A) 15  
B) 16  
C) 30  
D) 32  
E) 45