EFFECTS OF GENDER STEREOTYPES ON CHILDREN'S BELIEFS, INTERESTS, AND PERFORMANCE IN STEM FIELDS

By

Christine K. Shenouda

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Psychology—Doctor of Philosophy

2014

ABSTRACT

EFFECTS OF GENDER STEREOTYPES ON CHILDREN'S BELIEFS, INTERESTS, AND PERFORMANCE IN STEM FIELDS

By

Christine K. Shenouda

There is consistent evidence that women are underrepresented in science, technology, engineering, and mathematics (STEM) fields. Although boys and girls start out performing similarly on STEM-related school subjects, the gap between them widens as they mature, leading to the underrepresentation of women in STEM fields at the college level and in the workforce. One prominent interpretation of this widening gap is that women are conforming to gender stereotypes. Learned gender stereotypes affect people's interests, their beliefs in their abilities, and their performance on tasks. In this work, I explore two potential factors that may be contributing to the gender gap in STEM fields: children's susceptibility to stereotype threat and their beliefs about males' and females' competence and capacity to learn about STEM fields. In Study 1, two experiments investigated the effect of stereotype threat on children's (ages 4-9) performance on a spatial task. After a gender identity activation manipulation, children were asked to replicate a series of designs with LEGO blocks. In Experiment 1 ($N = 22$), preschool girls in the stereotype-threat condition were significantly slower than girls in the control condition, although they did not express strong explicit stereotypical attitudes in general, or specifically about LEGO blocks. In Experiment 2 ($N = 160$), children in kindergarten through third grade did not perform differently in the stereotype-threat condition, although girls in general were slower than boys. Moreover, children's speed on the block-construction task was predicted by their stereotypical attitudes towards LEGO blocks, whether their favorite toy was LEGO blocks, and how frequently they played with blocks at home. To complement these results, in Study 2 (*N* = 132), children rated males' and females' competence and learning ability in STEM and non-STEM fields. Children as young as age 5 held stereotypes about females' decreased ability to perform at STEM-related fields, although children's school-subject preferences and career aspirations did not reflect these stereotypes. Results from this work point to the detrimental effects of gender stereotypes on children's performance as early as age 4 and children's beliefs about women's abilities. Implications for understanding precursors of the STEM gender gap and for informing intervention work are discussed.

ACKNOWLEDGEMENTS

I would like to thank Judith Danovitch, my advisor and mentor, for her continuous support and guidance throughout this journey. Her flexibility and willingness to tap new research areas to fit my interests is unmatched. Her critical insights, thoughtfulness, and attention to detail has only made my work much better. Her positive outlook and how she calmly yet firmly handles situations has helped me learn to worry less and focus more on what needs to be done. Her enthusiasm for teaching and research and her dedication to her students have taught me a lot. I would also like to thank my committee members, Lauren Harris, Cary Roseth, and Isis Settles for all their helpful comments, guidance, and support. I am thankful to Nicholaus Noles for many insightful comments about this and other projects I have worked on. I would also like to thank professors Linn Liben and Rebecca Bigler for sharing their measures and materials and answering questions about them.

I am thankful to all the members of the Knowledge in Development (KID) Lab, especially undergraduate students who have helped me develop materials, run experiments, and enter data. I am especially thankful to Elizabeth Boudiab, Megan Fisher, Margaret Levasseur, Ashley Marderosian, and Jenna Shier.

I am grateful to all the children who participated in this research and every parent who took the time to read and sign a consent form and appreciated the importance of research. This research would not have been possible without the tremendous help of directors, staff, and teachers of numerous preschools and schools in Michigan and Illinois. I would like to especially thank Emily Weidner and Mary Vogt the principals of White Heath Elementary and Lincoln Elementary in Monticello, Illinois for allowing me into their schools and making data collection an enjoyable experience. This project would not have been possible without their appreciation for research and their flexibility.

I am grateful to my friend and commuting mate, Zornitsa Kalibatseva, for countless pep talks, informal therapy sessions, great insights about research and life, and many soul-warming meals. I would also like to thank Mai El-Hakim, my second commuting mate, who was there for me, rain, shine, or snow.

I am forever indebted to everyone in my family. My parents have always been there for me, supported me, and blessed my every move. The completion of this project would not have been possible without their countless sacrifices. I am thankful to my daughters who are my inspiration and my source of motivation. Finally, to my husband and best friend, Sameh Tawfick. Thank you for believing in me, for your critical insights, and for your unlimited support. Most of all, thank you for many years of selfless love.

"Now to Him who is able to do exceedingly abundantly above all that we ask or think, according to the power that works in us" Ephesians 3:20

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The underrepresentation of women in science, technology, engineering, and math (STEM; NSF, 2013) careers has been the focus of recent research attempting to explore its causes and suggest pathways to bridge this gap. From a developmental perspective, the development of certain cognitive abilities is required to excel at these fields (Spelke, 2005). In addition, this gender disparity has roots in men's and women's experiences since early childhood that help shape their attitudes and preferences about STEM fields (Fredricks & Eccles, 2002).

Little is known about precursors of the underrepresentation of women in STEM starting in the preschool and early elementary-school years. To fill this void, this dissertation examines two aspects that potentially affect women's choices and opportunities later in life: stereotype threat and attitudes towards women in STEM fields. In a series of experiments involving 212 children ages 4 through 9, I investigated the emergence and development of susceptibility to stereotype threat and children's learned beliefs and attitudes about females' competence and learning ability in STEM fields.

Chapter 1 presents a review of the relevant research literature. The chapter focuses on the presence of women in science and possible precursors for their underrepresentation, including gender bias and STEM preferences, with a focus on developmental trends. In this chapter, I also discuss the literature on stereotype threat and its effects on performance among adults and children as well as proposed mechanisms that may underlie the effects of stereotype threat.

Chapter 2 presents Study 1, which consists of two experiments involving preschool and elementary-school children. Experiment 1 examined preschool girls' susceptibility to stereotype threat on a block-construction task when gender was activated. It revealed that girls in the experimental condition performed significantly more slowly than girls in the control condition. Experiment 2 expanded the age range to older children (ages 5-9) and included a sample of boys. Children in the experimental conditions did not perform differently than children in the control condition. However, overall, girls were slower than boys on the block-construction task. In fact, speed on the block-construction task was predicted not only by age and gender, but also by children's explicit stereotype about blocks, how frequently they played with blocks, and whether LEGO blocks were their favorite toy.

Chapter 3 presents Study 2, which looks at a different aspect of precursors for women's underrepresentation in science: children's attitudes towards females in science. Children ages 5 to 9 rated men's and women's competence on a series of STEM and non-STEM professions as well as boys' and girls' learning ability on a series of STEM and non-STEM subjects. Children rated females' competence overall and learning ability as lower than males' on STEM fields but not on non-STEM fields, suggesting that STEM gender stereotypes develop early.

Chapter 4 presents an integrated discussion of the results of Study 1 and 2. The main findings of this work are discussed in light of existing literature with proposed follow-up studies and future directions and a discussion of the implications of this work.

CHAPTER 1

WOMEN IN SCIENCE:

STATUS, ATTITUDES, PREFERENCES, AND PERFORMANCE

Historically, women have always been underrepresented in STEM careers. Some have argued that this is because women are innately less apt at cognitive skills required to excel at STEM fields than men (e.g., Baron-Cohen, 2004, Benbow & Stanley, 1980; Summers, 2005). Others have opposed this view, citing the similar performance of males and females from infancy through maturity on basic cognitive tasks required to excel at science and math, such as learning about objects and numbers (see Spelke, 2005 for a review). The gender-representation gap in STEM fields has indeed narrowed over the past few decades, suggesting that innate differences are unlikely to be the cause as a few decades would be too short for any evolutionary effects to be evident (Handelsman et al., 2005). Still, although females are generally more likely to enroll in undergraduate studies (NSF, 2013), male undergraduates are more likely than females to major in physics and engineering (Xie & Shauman, 2003) and the doctoral degrees awarded to women in physical science and engineering account for only 25 and 15 percent, respectively, of total doctoral degrees awarded in these fields (Handelsman et al., 2005). The proportion of women at subsequent academic levels continues to shrink, resulting in females representing only 6 and 4 percent of full professors in physical science and engineering, respectively (Handelsman et al., 2005).

One approach to understanding the underrepresentation of women in STEM is to take a developmental perspective. It is important to examine the relationship between males' and females' performance in STEM fields beginning in childhood in order to understand when it starts to emerge and to investigate possible causes for a discrepancy, if one exists. Performance in math and science at schools has been the focus of research attempting to understand the origins of the STEM gap. With respect to math, despite the widespread contention that boys outperform girls at math, a meta-analysis of 100 studies of boys' and girls' math performance found that girls outperformed boys on computation in elementary and middle school, but the effect sizes were small[1] (Hyde & Linn, 2006; Hyde, Lindberg, Linn, Ellis, & Williams, 2008). Also, there was no difference between boys' and girls' performance on problem-solving tasks in elementary and middle school and only a small difference in favor of boys in high school. With respect to science scores as measured by standardized tests, boys outperform girls at all three schooling levels by an average of five points on a test out of 300 points (The Nation's Report Card, 2013). Although this difference is statistically significant, it is important to consider that these studies involve large samples and small effect sizes (Hyde & Linn, 2006). Thus, boys and girls start out performing similarly on math but the gap starts to show towards the end of schooling, whereas the science gap is consistent (albeit small) throughout the school years.

It seems unlikely that boys' and girls' discrepant innate cognitive abilities are to blame for the underrepresentation of women in STEM fields, given that they start out performing similarly. Why, then, do females underperform or lose interest in these fields during later school years leading up to an enormous gap in the workforce? Recent studies suggest one plausible explanation: attitude rather than aptitude (Else-Quest, Mineo, & Higgins, 2013; Shapiro & Williams, 2012; Cundiff, Vescio, Loken, & Lo, 2013). Attitudes about math and science can

[1] In general, effect sizes are often a more accurate measure of the magnitude of an effect, as large sample sizes can result in significant results even if differences are small (Hyde, 2005).

affect women's decisions to pursue careers in science by affecting personal preferences and by influencing the behavior of women's colleagues, bosses, and families.

Gender Bias in STEM

Cross-cultural research shows that stereotypes about science and math may be universal (e.g. del Rio & Strasses, 2013; Meltzoff & Kapur, 2014; Nosek, Banaji, & Greenwald, 2002a). There is strong evidence suggesting that males and females associate science with men (Nosek, Banaji, & Greenwald, 2002a). For example, sixty thousand American participants who completed an Implicit Associations Test[2] (IAT), showed a strong association between males and science (*d* = .72). Based on self-report questions, there was also a strong explicit association between males and science. Interestingly, implicit and explicit associations were stronger among older respondents, especially adults over fifty, suggesting that the narrowing gender gap in STEM may possibly be mirrored by less stereotypic attitudes among younger age groups.

Associating science with males also affects how people evaluate the competence of males and females in STEM fields. In a recent study, both male and female science faculty rated a student applying for a laboratory manager position as significantly more competent when the applicant had a male name than when the applicant had a female name (Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012). This suggests that stereotypes can taint judgments of ability and that both men and women are equally likely to commit such stereotypical judgments.

[2] A widely used implicit measure of stereotypes is the Implicit Association Test (IAT, Greenwald, McGhee, & Schwartz, 1998), which uses a series of trials to measure people's implicit associations about a variety of topics, such as associating science with males or aggression with Black people. By measuring how participants' speed in responding to stimuli is affected by pairing different topics together (e.g., math and females), scientists are able to assess participants' implicit associations of different topics.

Gender bias is not only manifested by adults in the workplace; early on, children are exposed to gender bias about math through their parents (Eccles & Jacobs 1986; Eccles et al., 1990; Jacobs & Eccles 1992; Midgley et al., 1989; Yee & Eccles 1988). Parents often believe that boys have higher math ability, and parents have higher expectations of math performance for boys than for girls, even though boys and girls perform similarly on tests. Further, parents' stereotyped beliefs affect their children's views of their own math ability (Jacobs, 1991; Parsons, Adler & Kaczala, 1982). Parsons et al. (1982) found that among fifth through eleventh graders, children's self-perceptions were affected more by their parents' beliefs in their abilities than by the children's own past performance. In addition, Jacobs (1991) found that among children in grades six through eleven, parents' gender stereotypes about math abilities affected their views of their child's abilities, which in turn affected children's self-perception and performance. Teachers also hold gender stereotypes about students' math abilities (see Gunderson, Ramirez, Levine, & Beilock, 2012 for a review). For instance, elementary-school teachers often believe that their male students like math more than female students (Fennema et al., 1990) and that math is more difficult for girls than for boys (Tiedemann, 2002).

Adults' biases also affect children's beliefs about the capabilities of males and females. In a recent study with Chilean kindergartners, children were asked to indicate which school subject (language or math) a boy and a girl liked more, which one they were better at and which one they liked better. Children expressed no preference when judging the male character but expressed that the female character would like language more, would be better at language, and would find math more difficult (del Rio & Strasses, 2013). These results suggest that children develop gender biases about school subjects by the time they enter elementary school.

Less is known about children's attitudes towards STEM careers. Children hold strong stereotypes about adult occupations. By age 5, children attribute better job performance to individuals who are presented in gender-stereotypical occupations (e.g., nurse, secretary, police officer, truck driver; Gettys & Cann, 1981). Similarly, preschoolers and first graders attribute more competence to a person of their own sex when the person is engaged in a gender-stereotypical occupation and they express more positive affective reactions when asked to imagine growing up to have a gender-stereotypical occupation (Levy, Sadovsky, & Troseth, 2000). Five- and six-year-olds are also more likely to misremember or distort information about individuals with counterstereotypical pairings of gender and occupation (Cordua, McGraw & Drabman, 1979; Wilbourn & Kee, 2010). In addition, Liben, Bigler, and Krogh (2001) found that children were more likely to identify with occupations that are typically held by individuals of their own gender. However, to my knowledge, no studies have looked at children's stereotypes about STEM and non-STEM professions. There is evidence that when a man or a woman are presented in a counterstereotypical profession, children still attribute knowledge about the profession to the proscribed professional, regardless of gender (Shenouda & Danovitch, 2013). For example, when children were presented with a male nurse and a female mechanic and asked who would know more about a series of profession-related items, children attributed mechanic-related knowledge to the woman and nurse-related knowledge to the man, even though their professions went against the stereotype. However, it is unclear whether children attribute more knowledge about the same profession to a male or female based on gender stereotypes.

STEM Ability and Preferences

Women express more negativity toward math than men, they identify less with math, and are less confident about doing math (Hyde, Fennema, Ryan, et al., 1990). This negativity, lower identification, and lack of confidence in math ability was more pronounced in high school and among undergraduate students than among younger students. Implicit measures also demonstrate a gender difference. In a study using the IAT, both men and women expressed negative implicit associations with math, but women had even more negative associations (Nosek, Banaji, & Greenwald, 2002b). By extension, women also identified less with math than men did. Identifying with math was assessed implicitly by measuring how fast participants responded when math was paired with self. In addition, women who identified more with being female showed weaker identification with math. Thus, females are more likely to have negative associations with math and are less likely to identify with math than males.

Learned gender stereotypes about science also affect how adults and children view their ability and their personal preferences. Cundiff et al. (2013) tested undergraduate students taking science classes to explore whether stereotypes predicted science identification and science aspirations using implicit and explicit measures. Women showed weaker science identification and science aspirations when they had strong gender-science stereotypes. On the other hand, men with stronger stereotypes showed stronger science identification and aspirations. Among children, several studies using self-report measures have demonstrated that girls rate their math abilities lower than boys (e.g., Fredericks & Eccles, 2002), and that this belief starts to emerge as early as the first grade (Entwistle, Alexander, Pallas, & Cardigan, 1987) and it exists even when there is no significant difference in actual performance (Herbert & Stipek, 2005). Children also exhibit similar trends on implicit measures. Cvencek, Meltzoff, & Greenwald (2011) used implicit as well as explicit measures to assess elementary-school children's attitudes towards math. They modified the IAT to be used with children, and used a math-gender stereotype IAT to measure the extent to which children associated math with their own gender. Girls ages 6-10 were less likely to associate math with their own gender, both on the IAT and on self-report measures. In addition, Cvencek, Meltzoff, and Kapur (2014) tested Singaporean elementary school children and found that, similar to Cundiff et al.'s (2013) results with adults, strong gender stereotypes and gender identity predicted identifying with math.

Taken together, research on math and science stereotypes suggests that children develop the stereotype that girls are not good at math and science before it has a direct impact on their performance. Thus, girls' knowledge of the stereotype may be driving performance gaps, rather than the other way around. In light of research indicating that undergraduates show lower interest in academic fields and occupations that do not match their academic self-concept (e.g., Frome, Alfred, Eccles & Barber, 2006), girls' underrepresentation in STEM fields could potentially be attributed to such stereotypes.

Effects of Gender Stereotypes on Performance

From a cognitive perspective, stereotypes are believed to be an adaptive tool that helps free up cognitive load to process more information (Bodenhausen & Wyer, 1985). For example, when subjects are asked to perform two tasks simultaneously, they perform better on one task when the other task has stereotypical heuristics to guide their performance (e.g., being asked to form impressions of people whose pictures are labeled with stereotypical labels, such as skinhead; Macrae, Milne, & Bodenhausen, 1994). Conversely, people are more likely to use stereotypes to judge others when their cognitive resources are depleted (Gilbert & Hixon, 1991), or when they are tired (Bodenhausen, 1990).

Although stereotypes can be a useful tool to free up cognitive resources, they have negative effects on one's beliefs about self and others and on performance. For example, making a social category salient by asking people to specify their gender before a math test or reminding participants of the stereotype that their social group lacks math ability has negative effects on females' performance, creating stereotype threat by activating the stereotype that females are weaker at math (e.g., Beilock, Rydell, & McConnell, 2007; Shih, Pittinsky, & Ambady, 1999). Stereotype threat is the phenomenon of decreased performance by a group of people when their group membership has been activated, provided that this group is stereotyped as lacking skill at the task at hand (see Aronson & McGlone, 2009 for review). This effect was first observed in a study by Steele and Aronson (1995) among African Americans, whose performance on a standardized verbal test was compromised by activating their race before the test and indicating that this test measured their intellectual ability. Many studies have since used similar methods with different categories, such as gender and ethnic identity, and have yielded similar results (see Aronson & McGlone, 2009 for review).

With respect to gender, prior to stereotype threat research, it was widely established that whereas women's and men's performance on easy math tests is comparable, women perform significantly worse on difficult or advanced math tests (see Spencer, Steele, & Quinn, 1999 for review). These results were either argued to be a product of genetic sex differences (Benbow & Stanley, 1980) or a result of socialized gender roles (e.g., Eccles, 1987), which encouraged males more than females to excel at math. Spencer et al. (1999) assessed women's performance on difficult math tests and found that when women were told upfront that the test did not yield differences between men and women, female test takers performed significantly better than women who were not given this description before the test. In fact, in this experimental condition, females' performance was comparable to males' performance on the difficult math test, controlling for SAT scores or equivalent. Together, these experimental studies suggest that stereotype threat affects performance on real-world tasks (such as math tests) and that there are ways to alleviate the negative effects of stereotype threat on performance (i.e., *stereotype lift*).

Although there is ample work demonstrating stereotype threat in adults, research on stereotype threat with children has been more limited. In one of the earliest studies with children, Ambady, Shih, Kim, and Pittinsky (2001) predicted that activating gender or ethnic identity in Asian American girls would result in differences in performance on a math test. In this study, Asian-American girls in kindergarten through eighth grade were randomly assigned to one of three conditions: female-identity activation, Asian-identity activation, and no activation. After a manipulation task, all children were given a standardized math test appropriate to their grade level. For the younger age group (kindergarten – grade 2) the manipulation task was to color a picture. A picture of a girl holding a doll was used to activate gender, a picture of two Asian children eating with chopsticks was used to activate ethnicity, and a landscape scene was used in the control condition. For the older age group (grade 3-8), children answered a series of questions about gender (e.g., whether boys or girls were better at school), or ethnicity (e.g., whether they spoke another language at home), or neutral questions (e.g., whether they liked animals). Overall, girls performed best at the math test when their ethnicity was activated and worst when their gender was activated. This is consistent with previous studies with adults (Shih et al., 1999) and with the stereotype that Asians are good at math and that females are not good at math. Thus, girls were highly susceptible to ST, and the activation of an identity associated positively with math helped them overcome this threat.

Building on this work, Neuville and Croizet (2007) investigated whether task difficulty affects performance under gender identity activation. French third-graders were first given the same gender activation manipulation (picture coloring) as in Ambady et al. (2001). Then, children solved seven arithmetical problems within a set time for each problem in a real classroom setting. There were five easy and two difficult problems. Compared to the control condition, girls' performance in the identity-activation condition was significantly better on easy problems and significantly worse on difficult problems. Boys' performance was not affected by problem difficulty. The stereotype threat effect demonstrated by girls' impaired performance on difficult problems replicates earlier findings (e.g., Ambady et al., 2001). However, girls' stronger performance on easy problems warrants an explanation. The authors interpret this results in terms of stereotype threat as a form of arousal (Neuville & Croizet, 2007). Arousal facilitates performance on easy tasks and hinders performance on difficult tasks (Brehm, 1999). Thus, when girls' gender identity was activated, they were more aroused and therefore performed better on easy problems but worse on difficult problems.

It could be argued that girls' stereotype level affects their susceptibility to stereotype threat, such that girls who hold stronger stereotypes are more vulnerable to stereotype threat. Ambady et al. (2001) tested this possibility by measuring children's explicit and implicit awareness of the math stereotype and its relationship to performance. To measure their implicit stereotype awareness, children heard an elaborate story about a student who was really good at math. The story did not refer to the child's gender. Then, participants were asked to repeat the story, and the experimenter noted whether they referred to the child in the story as "he" or "she."

To measure children's explicit stereotype awareness, children were asked: "Are boys better at math, girls better at math, or are they the same?" (similar tests were administered to measure ethnicity-stereotype awareness, but these are beyond the scope of the current discussion). Boys and girls were highly likely (75% of girls, 79% of boys) to indicate that both boys and girls are good at math when asked the question explicitly. However, girls were slightly (but not significantly) more likely than chance to infer that the child in the story was a boy, whereas boys were significantly more likely to indicate that the child was a boy (above 78% of boys). The authors concluded that stereotype threat is more closely related to implicit stereotypes than explicit stereotypes.

Another recent study (Galdi, Cadinu, & Tomasetto, in press) looked at the relationship between implicit associations and stereotype threat more systematically. In this study, first-grade girls showed susceptibility to stereotype threat when primed with the "girls are not good at math" stereotype by coloring a picture of a girl failing to solve a math problem correctly. Moreover, girls' susceptibility was mediated by their automatic association of male and math on an IAT, even when they did not show explicit math stereotypes. Thus, the authors suggest that stereotype awareness is not a prerequisite for stereotype threat. Together, these studies suggest that stereotype threat effects may be detected even among younger children who do not hold explicit stereotypes about gender.

How a task is presented to participants can also affect their performance. Huguet and Regner (2009) showed French children ages 11-13 a complex geometric figure and asked them to recreate it from memory. Children were instructed that this was either a task that tested their geometry ability or drawing ability. After completing the task, children were given a questionnaire to assess their level of stereotyping of geometry. Girls performed significantly worse in the “geometry” condition than in the “drawing” condition (even though it was the same task), whereas boys’ performance was similar in both conditions. On the stereotyping questionnaire, both boys and girls indicated that their own gender had superior geometry abilities, which suggests that even when girls hold explicit counterstereotypical beliefs, they are still susceptible to stereotype threat. It would have been interesting to measure children’s implicit stereotypes (as in Ambady et al., 2001), but the authors did not use an implicit measure in this study.

Overall, these studies reveal detrimental effects of stereotypes on girls’ performance on mathematical tasks, which are typically stereotyped as masculine. Such threat is lifted when the focus is diverted to another category with a positive stereotype in that domain (such as ethnicity, Ambady et al., 2001), or the task is labeled in a way that does not activate a stereotype (Huguet & Regner, 2009). Interestingly, children’s explicitly held stereotypes are not a determining factor in children’s susceptibility to stereotype threat. Rather, it is their implicitly held beliefs that are directly related to effects of stereotype threat on their performance (Ambady et al., 2001; Galdi et al., in press).

Within the limited research on stereotype threat in children, Ambady et al. (2001) is the only study that includes children as young as age 5. Given that research on stereotype threat with children has been focused on academic performance, particularly math tests, it may not have been of interest to researchers to study younger children, as they are less exposed to math and its stereotypes. However, measuring whether young children are susceptible to stereotype threat in domains where they have already acquired gender stereotypes, such as toy and play preferences, can address important questions about the nature and origins of stereotype threat. We know that children learn gender stereotypes very early on (e.g., Serbin, Poulin-Dubois, & Eichstedt, 2002), but have children internalized these stereotypes sufficiently that these stereotypes have negative effects on their performance? It is especially important to find out about these effects early in children's lives as children have many opportunities to learn and may potentially miss learning opportunities due to fear of failure as a result of learned stereotypes (Davies, Spencer, & Steele, 2005).

Stereotype Threat: Mechanisms

Research on stereotype threat has identified two main possible mechanisms by which the activation of negative stereotypes results in hindered performance: anxiety and depletion of working memory capacity (see Aronson & McGlone, 2009 for a review). One proposed mechanism for stereotype threat is that when a negative stereotype is activated, this creates a stressful situation for participants who try to disprove this stereotype. Evidence from studies that used simple and difficult tasks under stereotype-threat conditions support this proposal. Participants perform better on simple tasks and perform worse on difficult tasks under stereotype threat compared to participants in control conditions, both among adults (O'Brien & Crandall, 2003) and among children (Neuville & Croizet, 2007). Given that physiological arousal is known to contribute to improved performance on simple tasks and to hinder performance on difficult tasks (Yerkes & Dodson, 1908), these findings suggest that there is a heightened level of arousal involved in stereotype-threat conditions. In addition, using physiological measures, Blascovich, Spencer, Quinn, and Steele (2001) tested African American participants on a verbal task under control or stereotype-threat conditions (similar to Steele & Aronson, 1995) while continuously monitoring their blood pressure. In their study, participants in the stereotype-threat condition had heightened levels of blood pressure compared to African Americans in a control condition and compared to European Americans, and these heightened levels of blood pressure persisted even after the task was completed through a short break and a second unrelated task. However, using self-report measures of anxiety, participants did not indicate perceived higher levels of anxiety. Therefore, stereotype threat could be causing anxiety which affects both physiological and cognitive aspects, without participants being aware of either (Aronson & McGlone, 2009). This is possibly why stereotype threat studies that attempted to use self-report measures of anxiety yielded mixed results (e.g., Osborne, 2001; Spencer et al., 1999; Steele & Aronson, 1995, see Aronson & McGlone, 2009 for a review).

Another prominent proposal for the mechanism responsible for hindered performance under stereotype-threat conditions is compromised working memory capacity (Schmader & Johns, 2003). Working memory capacity refers to both attentional ability and temporary storage of information (Engle, Tuholski, Laughlin, & Conway, 1999). People with higher working memory capacity are better able to focus on a given task and ignore irrelevant information, and higher working memory capacity is related to superior performance on a variety of cognitive tasks (e.g., Rosen & Engle, 1998, see Schmader & Johns, 2003 for review).

Schmader and Johns (2003) explored how stereotype threat affects working memory capacity. They gave participants (males and females in Experiment 1, Latinos and Whites in Experiment 2) a dual-processing task where subjects had to judge whether an equation was right or wrong (e.g., "is (9 x 6) – 4 = 50?") while trying to memorize a word that appeared after each equation. After each block of trials, subjects were asked to recall as many words as they could, which was used as a measure of working memory capacity. In the stereotype-threat condition, where subjects were reminded of the gender stereotype (Experiment 1) or their racial identity was activated (Experiment 2), women and Latinos recalled significantly fewer words than males or White subjects respectively, unlike the control condition, where groups performed similarly.

In Experiment 3, Schmader and Johns gave female participants a standardized math test in addition to the working memory capacity task to assess whether working memory capacity mediates performance on the math test in the stereotype-threat condition. In this experiment, they modified the dual-processing task, by replacing the equation-judgment task with a task where participants had to count the vowels in words. This change was introduced to rule out an alternative explanation of the results from Experiments 1 and 2, that participants' performance on word recall was compromised because they were dedicating more attention to the equation-judging task as a result of the stereotype threat manipulation. Therefore, the authors replaced the equation-judging task with a vowel-counting task to make the task unrelated to the stereotype at hand: math ability. As predicted, women's performance on both working memory capacity and the math test was compromised in the stereotype-threat condition only. In addition, a mediational analysis revealed that working memory capacity was a significant predictor of math test performance (but not the other way around). Based on results from these three experiments, Schmader and Johns (2003) concluded that stereotype threat affects performance when the activated negative stereotype interferes with attentional resources.

Several other studies have come to similar conclusions. For example, Beilock et al. (2007) showed that when participants were allowed to practice hard math problems, thus reducing the need for working memory, their performance was not affected by stereotype threat. Inzlicht, McKay, and Aronson (2006) have also shown that stereotype threat negatively affects African Americans' performance on a Stroop task, which is a classic measure of cognitive interference. Jamieson and Harkins (2007) have further shown that stereotype threat affects women’s performance on an antisaccade task, another measure of inhibition and interference (see Schmader, Johns, & Forbes, 2008 for review).

In an effort to integrate these proposed mechanisms, Schmader et al. (2008) have proposed a process model that suggests that stress, active monitoring, and self-regulation consume cognitive resources and result in disrupted performance. In addition to the above-discussed physiological signs of stress, Schmader et al. proposed that stereotype threat drives individuals to continuously monitor the situation for cues that would for example overturn the relation between them and the group they belong to. Some studies support this notion by providing evidence that participants who identify with math tend to dissociate themselves from attributes associated with the group under stereotype threat (Pronin, Steele, & Ross, 2004; Steele & Aronson, 1995). Others have proposed that participants closely monitor their performance, as supported by studies showing that participants under stereotype threat work more systematically to avoid failure and worry more about their performance (Beilock et al., 2007; Seibt & Förster, 2004). The third mechanism that Schmader et al. (2008) propose is self-regulation: participants under stereotype threat exert effort to self-regulate their negative emotions about their group. Indeed, stereotype threat research has shown that participants show signs of such negative emotions as self-doubt and worry (e.g., Beilock et al., 2007; Steele and Aronson, 1995). The authors further propose that inconsistent results of self-report anxiety measures may be attributed to participants’ effortful attempts of self-regulation. Schmader et al. (2008) propose that these three mechanisms: physiological stress, monitoring, and self-regulation tax working memory capacity, resulting in depleted cognitive resources available for the task under study.

To my knowledge, no research has systematically examined the relationship between stereotype threat and working memory capacity in children. However, given that children’s working memory capacity is limited, and that it improves with age (Siegel & Ryan, 1989), I expected children's performance to be more compromised by stereotype threat in younger participants who also have limited working memory capacity.

The Current Study

In a series of experiments, I explored possible precursors of STEM underrepresentation by examining the emergence and development of stereotype threat and beliefs about competence and learning ability of STEM fields in children ages 4-9. This research addressed a number of questions. First, **when does gender-based stereotype threat start to emerge?** The study of stereotype threat with the youngest children, Ambady et al. (2001), grouped 5-year-olds with children up to second grade for the purpose of data analysis, and it is unclear how many 5-year-olds were in the sample (the authors did not provide means or standard deviations of age groups). By including preschoolers, the current study examines whether stereotype threat is exhibited in age groups younger than those that have been already studied. Preschoolers are old enough to have learned gender stereotypes (Ruble, Martin, & Berenbaum, 2006), yet the question remains whether these stereotypes already divert their attention and reduce their available working memory capacity and consequently affect their performance. Stereotype awareness is a prerequisite for stereotype threat (Aronson & Good, 2003). Therefore, if girls associate the task at hand with males, whether explicitly or implicitly, then one should expect girls, even as young as age 4, to exhibit stereotype threat when their gender is activated.

Second, this project examines whether stereotype threat is manifested in a domain that is not typically considered in an evaluative way. Much of stereotype threat research, especially with children, focuses on math ability. Math ability is typically stereotyped in the form of "boys are better than girls", which implies intelligence and skill. Stereotype threat effects follow clearly from this stereotype: girls are not as good as boys at math, therefore girls perform poorly when their identity is activated. However, toy preferences are not typically considered in a performance-related manner: boys are not perceived as being better at playing with blocks, they just prefer blocks.

Building with LEGO blocks is an excellent performance task for children because it is a quantifiable activity, and constructing a shape out of LEGO blocks based on a model requires significant working memory capacity. Furthermore, LEGO blocks are stereotypically associated with boys. In addition, block construction, as a spatial task is closely related to math (Verdine et al., 2014; Zhang et al., 2014, see also Mix & Cheng, 2012, for a review). In fact, a longitudinal study has found that preschoolers' block-building skills predict their math performance in middle and high school (Wolfgang et al., 2011). There is also a strong relationship between block-reproduction success and mental rotation skills (Brosnan, 1998; Jahoda, 1979). With respect to mental rotation, there is evidence that boys on average are better at mental rotation tasks than girls (e.g., Halpern et al., 2007), even in infancy (Moore & Johnson, 2008; Quinn & Liben, 2008) and among preschoolers (Hahn, Jansen, & Heil, 2010). In fact, spatial reasoning is one of the few cognitive skills that show consistent gender differences in meta-analytic studies (Hyde, 2005), and women show stereotype threat on mental rotation tasks (Moè, 2009; Moè & Pazzaglia, 2006). Recent research suggests that spatial training can improve math performance in children as young as age 6 (Cheng & Mix, 2014) and narrow the gender gap among undergraduates (Feng, Spence, & Pratt, 2007; Miller & Halpern, 2013). There is also recent evidence that among middle-school students, science test scores are mediated by mental rotation ability (Ganley, Vasilyeva, & Dulaney, in press), and some scientists argue that this could be beneficial for improving performance in STEM fields (e.g., Lubinksi, 2010; Newcombe, 2010). Thus, what starts out simply as playing with a toy may actually be a precursor for improved STEM-related skills and interest in STEM fields. I hypothesized that children would display stereotype threat even on this non-evaluative task, as it is typically associated with males.

Third, **how do children evaluate men's and women's competence and boys' and girls' learning ability on STEM and non-STEM fields?** By asking children to indicate how good men and women are at a variety of STEM and non-STEM professions and how well boys and girls can learn about new STEM and non-STEM topics, it is possible to identify whether children hold explicit STEM gender stereotypes. Based on similar research about children's stereotypes about math and language arts (del Rio & Strasses, 2013), I expected children to express stereotypical attitudes towards women in science by rating their competence and learning ability lower than that of male characters. I also expected older children to express more stereotypical attitudes, as they would have had more exposure and experience with these stereotypes.

Fourth, **what is the relationship between children's performance on the block-construction task and their attitudes and stereotypes about STEM, children's activities, and blocks**, as measured by self-report measures (Liben & Bigler, 2002) and an implicit measure (Ambady et al., 2001)? Similar to Galdi et al. (in press), I expected girls who hold stronger gender stereotypes to perform worse on the task when their gender is activated.

Finally, **are there gender and developmental differences on susceptibility to stereotype threat and on attitudes towards women in STEM-related fields**? With respect to stereotype threat, I expected girls to show susceptibility; however, boys' performance was harder to predict. In previous research, boys' performance on stereotyped tasks has been shown to either increase slightly (Ambady et al., 2001) or not change at all (Neuville & Croizet, 2007) when gender identity is activated. In addition, based on my previous research (Shenouda & Danovitch, in prep.), I expected children to exhibit in-group bias when rating males' and females' competence and learning ability. However, I also expected to see ratings follow stereotypical patterns. For boys, the STEM stereotype would be consistent with in-group bias (favoring males), thus boys were expected to show strong stereotypical patterns. Girls' responses, on the other hand, were harder to predict as in-group bias would compete with stereotypes. Whereas most research suggests that girls do show stereotypical patterns (e.g., del Rio & Strasse, 2013), other research has shown that girls show in-group bias on fields typically associated with males, at least on self-report measures (Huguet & Regner, 2009). Thus, I expected both in-group bias and STEM stereotypes to affect girls' responses, resulting in girls' responses being stereotypical for STEM fields but not as stereotypical as boys' responses. Developmentally, I expected older children to be more knowledgeable about STEM stereotypes, because children become more knowledgeable in general about stereotypes as they get older (Signorella, Bigler, & Liben, 1993). As for stereotype threat, I expected older children to be more knowledgeable about the blocks stereotype than younger children and thus to be more susceptible to stereotype threat. Chapters 2 and 3 address these research questions.

CHAPTER 2

STUDY 1: STEREOTYPE THREAT ON A BLOCK-CONSTRUCTION TASK

To investigate children's susceptibility to stereotype threat on a play-based task, children's gender identity was first activated following procedures from Ambady et al. (2001) and then children completed a task using LEGO blocks. The key question was whether girls in experimental conditions would be slower, less accurate, or both at completing the block-construction task than girls in the control condition. I was also interested in examining whether boys were in general faster and/or more accurate than girls and whether there were developmental differences in children's performance. In addition, I measured children's gender stereotyping levels in order to assess the extent to which their stereotypical beliefs related to their susceptibility to stereotype threat.

In Experiment 1, preschool girls colored either a gender-identity activating picture (a girl holding a doll) or a neutral picture (landscape) and then they completed a block-construction task, where they had to recreate as many of fifteen shapes of varying difficulty as they could in five minutes. In addition, girls' implicit (Ambady et al., 2001) and explicit (Liben & Bigler, 2002) stereotypes were assessed. In Experiment 2, boys and girls ages 5 to 9 completed a similar block-construction task preceded by a coloring task as in Experiment 1. Children also completed implicit and explicit stereotyping measures.

Preschool girls were included in this sample to investigate whether stereotype threat is observable in such a young age group, since no stereotype threat research to date has included children younger than age 5. Block construction was identified as a suitable task for this study for its stereotypical association, spatial nature, and its close relation to math. LEGO blocks are typically associated with males and preschool boys spend more time playing with blocks than girls (Goble, Martin, Hanish & Fabes, 2012). In addition, block construction constitutes a spatial task which requires working memory, and block construction has been linked to math ability (Verdine et al., 2014; Zhang et al. 2014). Thus, children's block construction ability may be a precursor to their later performance in STEM fields.

Experiment 1

Method

Participants

Twenty-two preschool girls (*M* = 4.61, *SD* = 0.31, range: 4.03-5.15) were recruited from two mid-sized Midwestern cities through local preschools and a laboratory database. Children were randomly assigned to either the experimental condition (*n* = 11, *M* = 4.56, *SD* = 0.39) or the control condition (*n* = 11, *M* = 4.66, *SD* = 0.22). The majority of children were Caucasian and non-Hispanic.

Procedure and Materials

All children were tested individually in a quiet room at their schools or in a laboratory. Sessions lasted approximately fifteen to twenty minutes. All children were interviewed by a female experimenter. Children were randomly assigned to either the experimental or control condition and the experimenter was blind to the condition. The experiment consisted of three procedures in the following order: gender-identity activation, block-construction task, and assessment of stereotyping.

Gender-Identity Activation: In the first portion of the experiment, children were asked to color a picture using crayons. Similar to Ambady et al. (2001), girls in the experimental condition colored a picture of a girl holding a doll and girls in the control condition colored a picture of trees (see Appendix A). Children were given the picture in an opaque folder randomly drawn from a set of folders so that the experimenter remained blind to the condition. Children were told that they would first color this picture and then they would “play a hard game with some blue blocks” and they had three minutes to color the picture (see Appendix B for full script).

Block-Construction Task: After three minutes, the experimenter asked each child to put the picture back in the folder and then introduced the block task. The experimenter first gave each child a practice shape to complete (see Appendix C) by putting a shape in front of children and giving them corresponding pieces to use to recreate the same shape. Children were corrected if they did not recreate the shape accurately. Then, the experimenter explained that the child would have five minutes to complete as many shapes as possible and that they should work as fast as they could but also make sure to get the shapes correct. Children were advised that they would not be corrected (in order not to reinforce their performance) and that they should indicate when they have completed a shape in order to move on to the next one. The experimenter then placed the first shape and its corresponding pieces in front of the child and started a five-minute timer. There were a total of 15 possible shapes to complete containing two to seven pieces and they became progressively more difficult (see Appendix C). These shapes were pretested with three children ($M = 4.61$, 2 boys), to ensure that 4-year-olds were able to recreate the shapes and that there was a progressive degree of difficulty. Every time children indicated that they had completed a shape, the experimenter put the next shape and corresponding pieces in front of them, took away the completed shape, and marked whether the child matched the target shape correctly. When time was up, the experimenter stopped the child and noted which shape the child was working on. Children were asked if they liked the game.

Assessment of Stereotyping: After completing the LEGO task, children completed an implicit stereotype measure (based on Ambady et al., 2001), the activities attitude subset of the Preschool Occupations, Activities, and Traits scale (POAT-AM, Liben & Bigler, 2002) and an explicit measure about LEGO blocks. To gauge whether children implicitly believe that boys are more skilled at LEGO blocks than girls, children heard the following story and were asked to repeat it: "There were many friends that were good at this block game, but one of them stood out from the rest. This friend worked very fast with the blocks, got the shapes all right and even put together new shapes that even I couldn't do. One time, this friend entered a block game competition and came in first place." The experimenter noted whether children referred to the child in the story using a masculine, feminine, or neutral pronoun.

To measure children's overall explicit attitudes towards toys, the activities subset of the PAOT-AM was administered (see Appendix E for POAT-AM list of items). The POAT-AM measure comprises fourteen items that represent toys and activities that are typically masculine (6 items, e.g., robots), typically feminine (6 items, e.g., dolls), or neutral (2 items, e.g., big bouncy ball). Children were presented with three pictures with stick figures that represent boys only, girls only, or both boys and girls and were trained to use this scale using two irrelevant items ("who should wash their hands before eating" and "who should wear a ribbon in their hair"). Then, children were presented with a picture of each of the 14 activities and toys and were asked to indicate who they thought should do these activities: only boys, only girls, or both boys and girls (see Liben & Bigler, 2002 for information about scale reliability and validity).

To test whether children have an explicit stereotype about who should play with LEGO blocks, using the same scale used in the POAT-AM measure, children were shown three pictures of LEGO blocks: one of multicolored blocks, one of blue blocks only, and one of pink blocks only and were asked to indicate who should play with these LEGO blocks, without the experimenter labeling the pictures.

Results and Discussion

Block-Construction Task

Two measures were devised to analyze children's performance on the block-construction task: a speed measure (how many shapes the child attempted) and an accuracy measure (how many shapes the child completed accurately as a proportion of total shapes attempted). An independent-samples *t*-test comparing children's speed revealed a significant effect of condition, $t(20) = 3.77$, $p = .001$, $d = 1.61$, such that children in the experimental condition completed fewer shapes than children in the control condition (experimental: $M = 8.18$, $SD = 1.94$; control: $M = 11.00$, $SD = 1.55$; see Figure 1). Looking at accuracy (proportion of shapes completed correctly), children in both conditions performed similarly, $t(20) = 0.34$, $p = .74$ (experimental: $M = .70$, $SD = 0.097$; control: $M = .72$, $SD = 0.12$). Thus, on the LEGO task, girls whose gender identity was activated showed stereotype threat by performing significantly slower. It is possible that girls were also threatened by the description of the study (having to work very fast and get shapes right) so that they worked more slowly, making sure to check their work and guarantee its accuracy. When asked if they liked the game, three children refused to answer and all others said they liked the game.

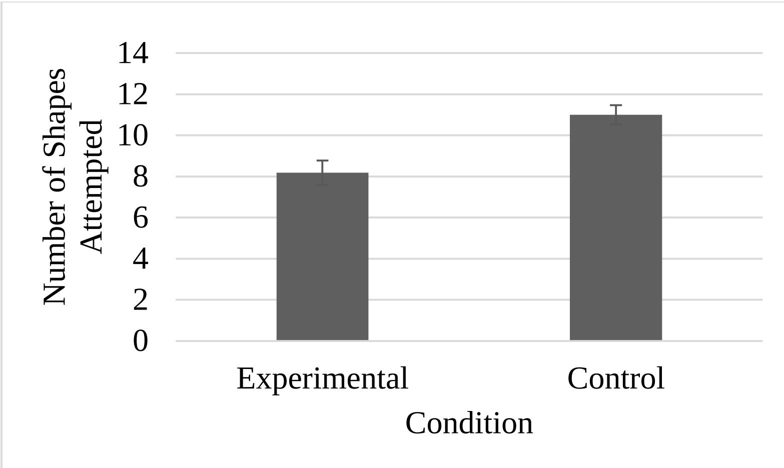

Figure 1. Experiment 1: Block-Construction Performance. Mean number of shapes attempted by condition. Error bars denote standard error of the mean.

Assessment of Stereotyping

To analyze children's explicit stereotypes, as measured by the activities subset of the POAT-AM, two composites were created, one for masculine items and one for feminine items, such that for each item children received one point for indicating a stereotypical answer and zero points for other answers (Liben & Bigler, 2002). A repeated-measures ANOVA with POAT Stereotype Category (female, male) as the within-subjects variable and Condition (experimental, control) as the between-subjects variable did not reveal a main effect of Stereotype Category or Condition (Fs ≤ 1.52, ps $\geq .23$). Thus, children's answers were similar across conditions and on masculine and feminine items. In fact, children's answers were not significantly different from chance on either masculine or feminine items (ts ≤ 1.65, ps $\geq .11$). In this case, chance was the likelihood of choosing the stereotypical answer (boys only for masculine items, girls only for feminine items) versus choosing either the counterstereotypical answer or choosing "both".

When children retold the story of the block-competition winner, overall children were more likely to refer to the child using a masculine pronoun (59%) than using a feminine pronoun

(14%) or referring to the child neutrally (27%), $\chi^2(2, N = 22) = 7.18, p = .028$. Seventy-three percent of girls in the control condition referred to the child using a masculine pronoun, whereas only 45% of girls in the experimental condition used a masculine pronoun when referring to the child. However, these results were not significantly different from each other (Fisher's Exact test, $p = 0.4$, Freeman & Halton, 1951).

To analyze children's explicit stereotypes about blocks, children received one point for every time they gave a stereotypical answer (boys only for multicolored and blue blocks, girls only for pink blocks) and zero points for all other answers (Figure 2). When girls were asked who should play with multicolored, blue-only, and pink-only blocks, overall, girls were more likely than chance to answer with "both boys and girls" or with "girls only" than they were to answer with "boys only" for multi-colored blocks ($\chi^2(1, N = 22) = 8.91, p = .003$) and blue-only blocks ($\chi^2(1, N = 22) = 6.55, p = .011$), whereas they were more likely than chance to answer with "girls only" for pink-only blocks ($\chi^2(1, N = 22) = 11.64, p = .001$). There were no effects of condition in either of the three questions (Fisher's Exact tests, $ps \geq .21$).

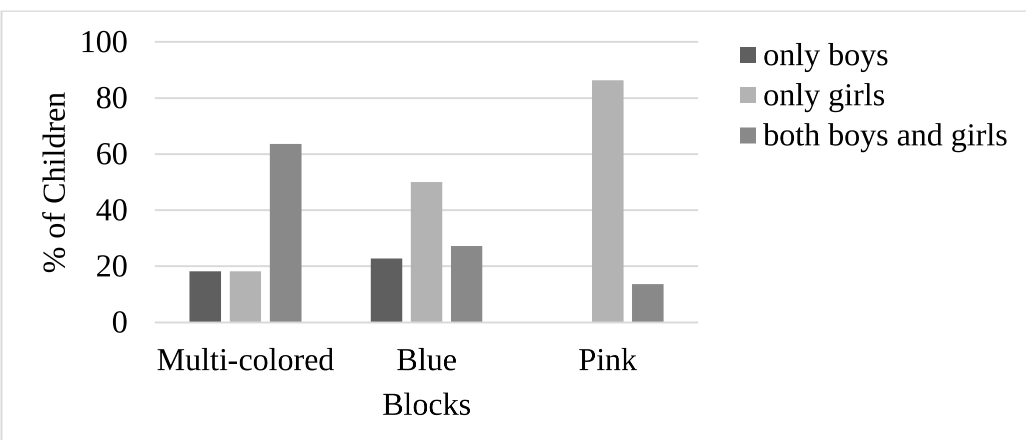

Figure 2. Experiment 1: Who Should Play with Blocks? Children's responses to the questions of who should play with multicolored, blue, and pink blocks.

Because children's answers to these three questions were likely affected by the block construction task that participants had just completed using blue blocks, I sampled an additional group of children from the same community in order to gather information about baseline responses (N = 27, M_{age} = 4.46, SD = 0.24, range: 3.93-4.90, 11 girls). These children only answered the three questions about the different-colored blocks (Figure 3). Looking at boys' and girls' responses separately, boys' and girls' answers were not significantly different from chance (which is the likelihood of choosing the stereotypical answer versus choosing the other two answers) on all three questions ($ts \leq 1.61$, $ps \geq .14$), except for girls' answers on multicolored blocks ($t(10) = -2.61$, $p = .026$). Girls were significantly less likely than chance to indicate that multicolored blocks are for boys only. Thus, even though boys are generally believed to spend more time playing with blocks than girls (Goble et al., 2012), and although girls did show slower performance in the experimental condition, our results do not show that preschool children strongly associate playing with blocks as being a masculine activity, at least explicitly.

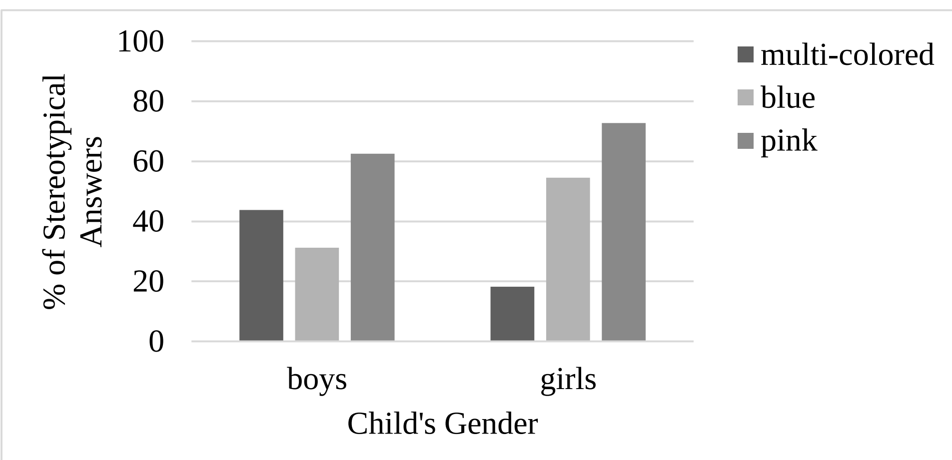

Figure 3. Control Sample: Who Should Play with Blocks? Twenty-seven boys and girls indicated whether multicolored, blue, and pink blocks are for boys, girls, or both. Percentages indicate stereotypical answers (boys only for multi-colored and blue blocks and girls only for pink blocks).

Taken together, the results of Experiment 1 suggest that preschool girls are susceptible to stereotype threat even when engaging in an activity that is not typically considered in an evaluative manner. Girls in the experimental condition were significantly slower than girls in the control condition, suggesting that the gender-identity activation procedure affected their performance by slowing them down. One limitation to the gender-identity activation task is that having girls color a picture of a girl holding a doll may have activated not just their gender identity but also toy stereotypes, thus inflating the stereotype threat effect. To address this possibility, it is necessary to include a condition where only gender identity is activated.

Although girls did not show strong stereotypical beliefs on the POAT-AM scale or explicit questions about blocks, they seemed to implicitly associate superior block competence with males. The discrepancy between children's implicit and explicit answers is consistent with similar findings in the literature (Ambady et al., 2001; Galdi et al., in press). However, it may be further explained by the slightly different question on the implicit and explicit block stereotyping measures. The implicit measure implied competence (a block-competition winner) whereas the explicit measure merely asked about who should play with these blocks, thus not necessarily tapping into competence. It is possible that children believe that both boys and girls should play with blocks but that boys are better at building with blocks. This issue will be addressed in Experiment 2.

Experiment 2

Experiment 1 showed that girls as young as age 4 are susceptible to stereotype threat, even when they do not hold explicit stereotypes about masculine and feminine activities in general and about blocks in particular. Some questions remain unanswered after Experiment 1.

First, how does this effect develop with age? Would older girls show stronger or weaker effects of stereotype threat on a similar task? Second, how would a gender-identity activation task affect boys' performance on a LEGO task? To answer these questions, Experiment 2 included boys and girls in kindergarten through third grade. As mentioned in Chapter 1, I expected older children to be more knowledgeable about the blocks stereotype than younger children and thus to be more susceptible to stereotype threat. Boys were included to investigate how their performance is affected by gender-identity activation and whether there are overall gender differences on the block-construction task, regardless of condition. Because previous research has not found consistent effects of stereotype threat on boys' performance (Ambady et al., 2001; Neuville & Croizet, 2007), no prediction about boys' performance was made.

In addition, Experiment 2 included a second experimental condition that did not include a stereotypical toy to examine whether girls perform differently when only gender is activated versus when gender and toy stereotypes are activated. The design was similar to Experiment 1, with a number of changes. First, the dependent variable in the block task was the time in which children completed eight shapes. Because in Experiment 1 the main effect of stereotype threat was on speed rather than accuracy, Experiment 2 used a more precise measure of speed rather than the proxy of how many shapes were completed. Second, to ensure that children processed the picture they colored sufficiently, at the end of Experiment 2 children were asked to indicate which picture they colored. Third, in addition to the POAT-AM, children completed the personal activities measure (POAT-PM) to assess their personal preferences for gendered toys and activities. This measure was added to assess whether children's personal preferences, not just their stereotypes, were related to their performance on the block-construction task. Although the POAT measure is primarily intended for preschoolers, I used it because I was interested in children's preferences and attitudes for this specific set of toys and activities. The measure intended for older children (COAT, Liben & Bigler, 2002) does not include these items and hence would not have allowed for comparison across age groups and between Experiments 1 and 2, on the POAT-AM measure. In addition, parents completed a brief questionnaire about their children's toy preferences in order to understand the relationship between toy preferences and performance.

Method

Participants

Children (*N* = 160, 78 girls)[3] were recruited from two small Midwestern cities through two public schools and a private school. Children were divided into two age groups: Children in kindergarten and first grade (*n* = 76, *M* = 6.33, *SD* = 0.58, range: 5.15-7.26, 37 girls) and children in second and third grade (*n* = 82, *M* = 8.25, *SD* = 0.64, range: 7.18-10.00, 41 girls)[4]. Children were randomly assigned to either of the two experimental conditions or the control condition (Table 1 includes cell sample sizes, means, and standard deviations). There were no statistical differences in age between the different conditions. One child was identified by his parents as Hispanic. Most children were identified by parents as Caucasian (89%), with one Asian child, three African-American children, two American Indians/Alaska Natives, four

---

[3] A total sample size of 64 children per gender, divided into two age groups, would be appropriate to detect a large effect size (d = 0.4).

[4] There was an overlap between the two age groups, such that the two youngest children in 2nd grade were younger than the oldest children in 1st grade. To avoid this overlap, another way of looking at this data would have been by age instead of grade. However, research shows that similarly aged children have a higher IQ if they have completed more years of schooling (see Nisbett, 2009 for a review). Thus, I split the data by grade instead of age.

multiracial children, and the rest unidentified.

	Boys					
	Landscape		Boy Only		Boy with Truck	
	n	M (SD)	n	M (SD)	n	M (SD)
K&1	15	6.46 (0.55)	13	6.38 (0.57)	13	6.16 (0.72)
2&3	14	8.17 (0.78)	14	8.23 (0.64)	13	8.16 (0.52)

	Girls					
	Landscape		Girl Only		Girl with Doll	
	n	M (SD)	n	M (SD)	n	M (SD)
K&1	12	6.49 (0.43)	13	6.15 (0.61)	12	6.30 (0.62)
2&3	12	8.34 (0.50)	14	8.33 (0.75)	15	8.26 (0.65)

Table 1. Experiment 2: Sample Statistics. Sample size, means and standard deviations of ages of boys and girls in both age groups by condition.

Procedure and Materials

All children were tested individually in a quiet room at their schools. Sessions lasted approximately fifteen to twenty minutes. All children were interviewed by the same female experimenter. Children were randomly assigned to one of the three conditions corresponding to their gender and the experimenter was blind to their condition. The experiment consisted of three main segments in the following order: gender-identity activation, block-construction task, and assessment of stereotyping. In addition, on the consent form, there was a brief questionnaire that parents were requested to fill out (Appendix F).

Gender-Identity Activation: In the first portion of the experiment, children were asked to color a picture using crayons. Girls colored either a picture of trees, a standing girl, or a girl holding a doll. Boys colored either the same picture of trees, a standing boy, or a boy holding a truck (see Appendix A). Children were given the picture in an opaque folder so the experimenter remained blind to the condition. Children were told that they would first color this picture and then they would "play a hard game with some blue blocks" and they had three minutes to color the picture (see Appendix D for full script). When children completed all segments of the experiment, they were given a strip of paper with thumbnails of all three possible coloring pictures (depending on child's gender) and were asked to circle the picture they colored, to guarantee that they paid attention to the picture and recalled it. The experimenter did not look as they circled the picture to remain blind to the condition. Two boys in the younger age group (in the boy-only coloring page condition) failed to circle the correct picture, and were eliminated from further analyses.

Block-Construction Task: After three minutes, the experimenter asked children to put the picture back in the folder and then introduced the block task. Similar to Experiment 1, the experimenter first gave children a practice shape to complete (see Appendix C). Children were corrected if they did not recreate the shape correctly. Then, the experimenter explained that children would have to complete eight shapes (see Appendix C) as fast as they could and to make sure to recreate the shapes accurately. They were also instructed to shout "done" when they were done so the experimenter could stop the timer. Children had to repeat the instructions to guarantee that they understood the task. The experimenter then placed the first shape and its corresponding pieces in front of the child and started the timer. Each shape and its corresponding pieces was placed on a white paper plate and children pulled each plate towards them in order to complete each shape. When children completed all eight shapes and shouted "done," the experimenter stopped the timer and noted their speed. At the end of the session the experimenter also noted whether children completed each shape accurately. After children completed the LEGO task, they were asked whether they liked the game. They were also asked whether the game was easy or hard, and depending on their answer, they were asked whether it was very easy/hard or somewhat easy/hard, creating a four-point scale.

Assessment of Stereotyping: After completing the blocks task, children completed the implicit stereotype measure from Experiment 1, the activities personal and attitude subsets of the Preschool Occupations, Activities, and Traits scale (POAT-AM and POAT-PM, Liben & Bigler, 2002) and an explicit measure about LEGO blocks.

The POAT-PM measure comprises 14 items, most of which overlap with the POAT-AM measure. Administration followed the procedures described by Liben and Bigler (2002). Children were first introduced to three faces and were trained to use them to express how much they liked ice cream ("a lot", "some", "not at all"; see Appendix E for POAT-PM list of items). Then, children were presented with a picture of each of the toys/activities and were asked to indicate how much they like to play with this toy using the three faces. Children were always administered the personal measure prior to the attitude measure, in order for preferences to be unaffected by gender stereotypes.

Unlike Experiment 1 where children were asked who should play with LEGO blocks, in Experiment 2, using the same scale used in the POAT-AM measure, children were shown the same three pictures of LEGO blocks (multicolored blocks, blue blocks only, and pink blocks only) and were asked "who is better at building with LEGO blocks" instead, in order to measure whether children had an evaluative attitude about LEGO blocks.

Parent Questionnaire: To collect more information about the children in the sample, on the consent form signed by parents, in addition to demographic information, parents answered 11 questions about their child. Specifically, parents were asked how many children live in their home and each child's age and gender. Parents also indicated the favorite toy of the child in the study and whether their child had access to any of the following toys at home and, if so, how often the child played with these toys (never, rarely, sometimes, frequently, all the time): LEGO blocks, dolls, cars/trucks, and a kitchen or tea set (see Appendix F for full set of questions).

Results and Discussion

Block-Construction Task

To analyze children's performance on the LEGO task, first, a 2 (Child's Gender) X 2 (Age Group) ANOVA was conducted to look at overall age and gender differences in speed. The ANOVA revealed a main effect of Age Group, $F(1, 154) = 35.68$, $p < .001$, $\eta_p^2 = .19$, such that children in the older age group were faster than children in the younger age group (K&1: *M* = 201s, *SD* = 75s; 2&3: *M* = 145s, *SD* = 45s). There was also a main effect of Child's Gender, $F(1, 154) = 8.69$, $p = .004$, $\eta_p^2 = .05$, such that boys were faster overall than girls (boys: *M* = 159s, *SD* = 61s; girls: *M* = 185s, *SD* = 71s). In addition, the Age Group x Child's Gender interaction was marginally significant, $F(1, 154) = 3.82$, $p = .052$, $\eta_p^2 = .02$, such that the speed difference between boys and girls in the younger age group was more pronounced than in the older age group (Figure 4). Removing outliers (outside the inner fences) and extreme outliers (outside the outer fences, Tukey, 1977) did not result in changes in result patterns. Thus, all subjects were included in this and further analyses.

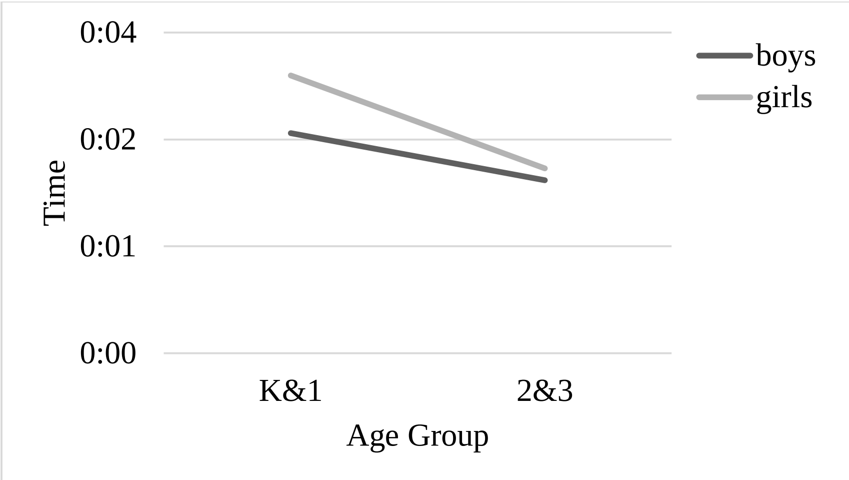

Figure 4. Experiment 2: Block-Construction Speed. Children's speed on the LEGO task by gender and age group.

To analyze the effect of condition, boys' and girls' speed was analyzed separately using two 3 (Condition) X 2 (Age Group) ANOVAs. Girls' performance did not show an effect of Condition, $F(1, 72) = 0.58$, $p = .56$, $\eta_p^2 = .016$ (Figure 5). There was a significant main effect of Age Group, $F(1, 72) = 30.77$, $p < .001$, $\eta_p^2 = .30$, such that older girls were faster than younger girls. Thus, an effect of stereotype threat among elementary-school girls was not apparent on the block-construction task. There was also no significant interaction between condition and age, $F(1, 72) = 2.12$, $p = .13$, $\eta_p^2 = .06$.

Looking at boys' speed on the block-construction task (Figure 6), there was a significant main effect of Age Group, $F(1, 74) = 9.15$, $p = .003$, $\eta_p^2 = .11$, such that older children were faster. Although there was no significant main effect of condition, $F(2, 74) = 0.32$, $p = .73$, $\eta_p^2 = .08$, there was a significant Age Group X Condition interaction, $F(2, 74) = 3.62$, $p = .032$, $\eta_p^2 = .089$. However, breaking down the interaction did not reveal any significant effects. Thus, boys' performed similarly regardless of the gender-activation condition.

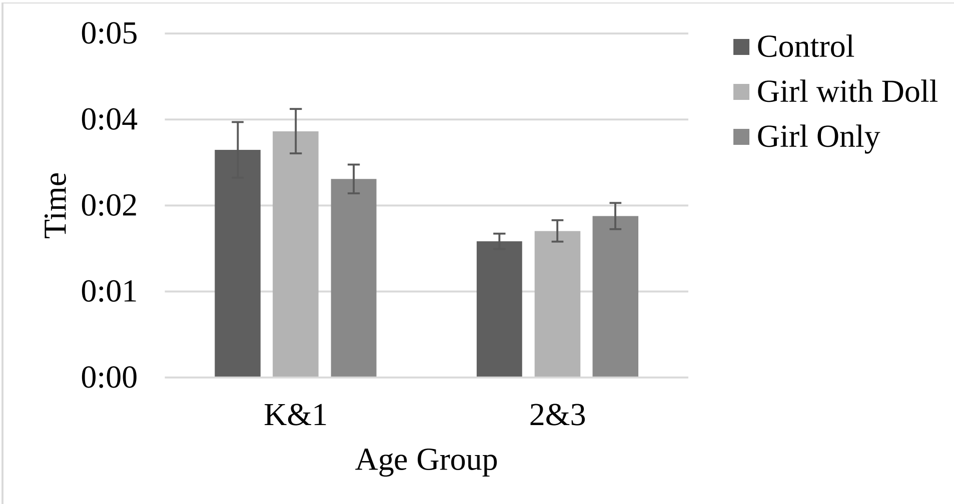

Figure 5. Experiment 2: Girls' Speed. Girls' speed on the block-construction task by age group and condition. Error bars denote standard error of the mean.

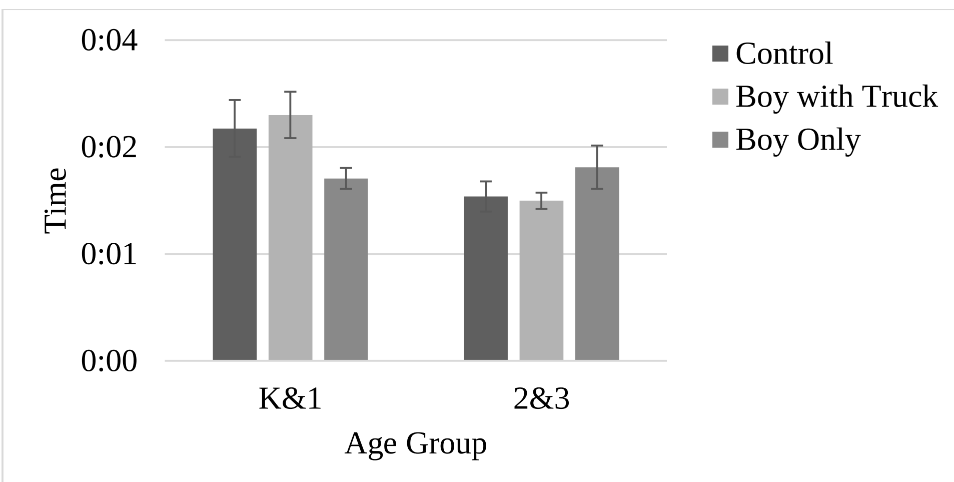

Figure 6. Experiment 2: Boys' Speed. Boys' speed on the block-construction task by age group and condition. Error bars denote standard error of the mean.

Children's accuracy (proportion of correct shapes completed) was also analyzed to explore whether there were any age, gender, or condition effects. A 2 (Child's Gender) x 2 (Age Group) ANOVA revealed a significant effect of Age Group, $F(1, 154) = 7.62$, $p = .006$, $\eta_p^2 = .047$, such that older children had higher accuracy levels. However, there was no significant effect of Child's Gender, $F(1, 154) = 0.87$, $p = .35$, $\eta_p^2 = .006$. Analyzing accuracy by condition also revealed no significant effects for either gender (Fs ≤ 1.63, ps $\geq .23$). Thus, whereas children became more accurate as they got older, unlike the speed measure, both boys and girls performed similarly with respect to accuracy. For comparison, Figure 7 shows accuracy rates from Experiment 1 and 2. Children became progressively more accurate as they got older.

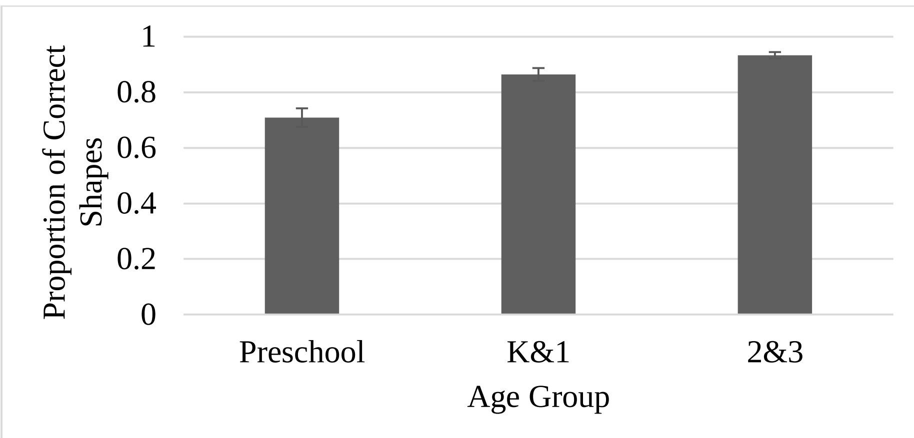

Figure 7. Accuracy Rates. Proportion of Shapes completed correctly across Experiments 1 and 2. Error bars denote standard error of the mean.

All children said they liked the block-construction game, except one boy who said he did not like it because "it was over too soon." When asked whether the game was easy or hard (Figure 8, Figure 9), most children indicated that it was very easy or somewhat easy, even though the experimenter introduced the task as a "hard game". A 2 (Child's Gender) X 2 (Age Group) ANOVA looking at children's answers to the difficulty question revealed a significant main effect of gender, $F(1, 154) = 9.49$, $p = .002$, $\eta_p^2 = .06$, such that girls rated the game as harder than boys. There was no significant main effect of Age Group, $F(1, 154) = 0.84$, $p = .36$, $\eta_p^2 = .005$. Looking at condition for each gender separately revealed no significant main effect of Condition for children's rating of difficulty (*F*s ≤ 2.47, *p*s ≥ .091). Thus, whereas all children liked the game, girls were more likely to indicate that they found the game to be hard. In fact, there was a significant moderate correlation between children's speed and their difficulty rating, *r*(158) = .229, *p* = .001, such that children who took longer to complete the task had a higher difficulty rating for the task.

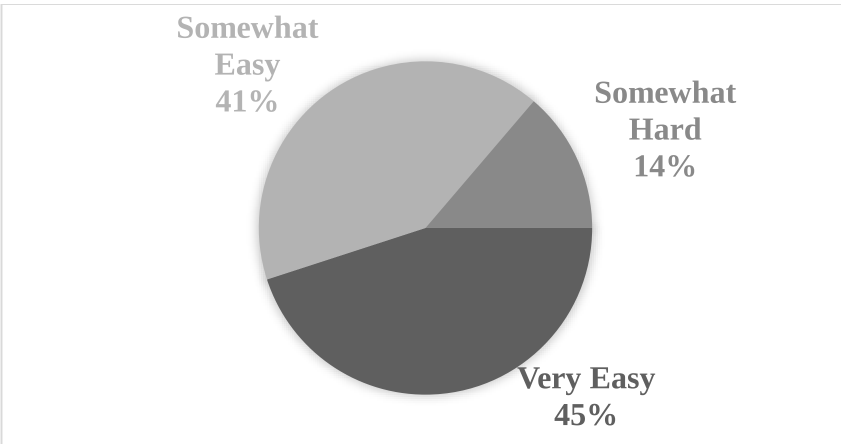

Figure 8. Boys' Difficulty Ratings. Boys' difficulty rating of LEGO task.

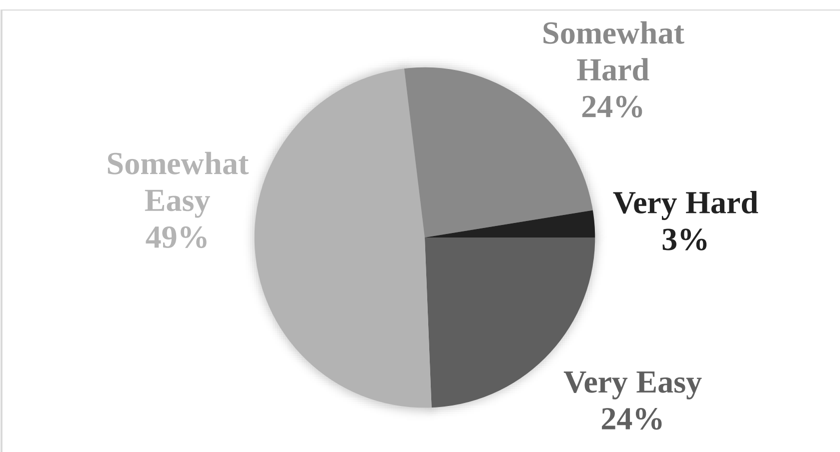

Figure 9. Girls' Difficulty Ratings. Girls' difficulty rating of LEGO task.

Stereotype Assessment

Children's responses to the POAT measure were analyzed using two repeated-measures ANOVAs to analyze the attitude measure and personal measure separately. Initial analyses revealed no significant effects of condition on children's stereotyping, and thus only Child's Gender and Age Group were included in the following analyses. With respect to the Attitude Measure (feminine and masculine), there was a significant effect of Measure, $F(1, 154) = 10.22$, $p = .002$, $\eta_p^2 = .06$, such that overall children were more stereotypical on feminine items than on masculine items. There was also a significant main effect of Child's Gender, $F(1, 154) = 4.93$, $p = .028$, $\eta_p^2 = .031$, such that boys were more stereotypical than girls. A significant main effect of Age Group was also detected, $F(1, 154) = 4.19$, $p = .042$, $\eta_p^2 = .027$ indicating that children in the older and younger age group had different response patterns. A significant Measure X Age Group interaction further explained this effect, $F(1, 154) = 12.84$, $p < .001$, $\eta_p^2 = .077$, showing that children in the younger age group were more stereotypical on masculine items than children in the older age group ($t(156) = 3.36$, $p = .001$) and that their responses were equally stereotypical on feminine items ($t(156) = 0.93$, $p = .35$). Overall, children's responses were also significantly different from chance, $t\text{s} \geq 7.72$, $p\text{s} < .001$. Taken together, children's opinions about activities and toys (POAT-AM) suggest that they have gender stereotypes about masculine and feminine activities, that younger children are more stereotypical about masculine items, and that boys overall are more stereotypical than girls (Figure 10, Figure 11).

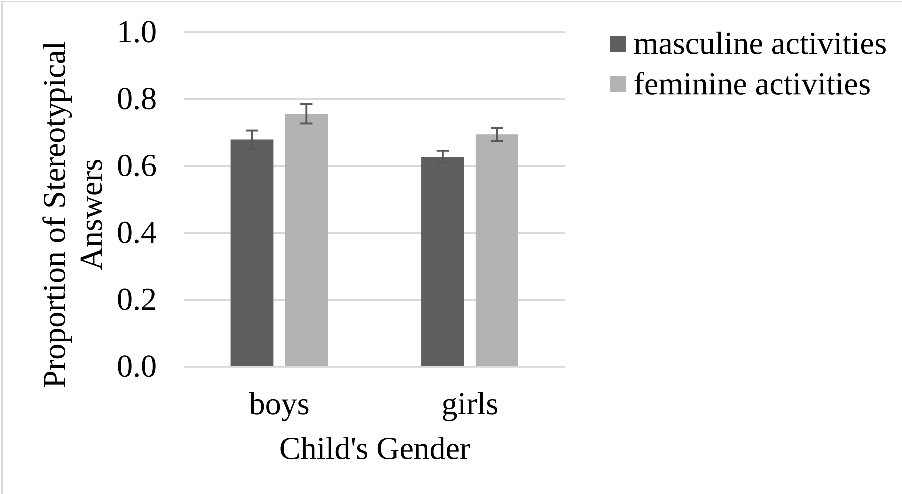

Figure 10. POAT-AM by Gender. Proportion of stereotypical answers by gender on Activities subscale. Error bars denote standard error of the mean.

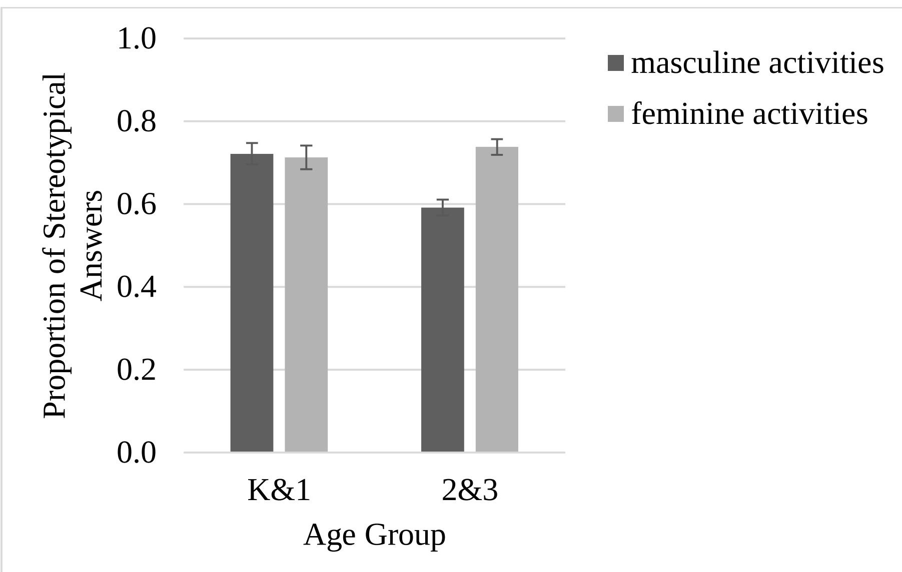

Figure 11. POAT-AM by Age Group. Proportion of stereotypical answers by age group on Activities subscale.

Looking at children's personal preferences for masculine and feminine items, there was a significant main effect of score on the Personal Measure (feminine, masculine), $F(1, 139) = 91.14$, $p < .001$, $\eta_p^2 = .396$, such that overall children preferred masculine items over feminine items. There was also a significant main effect of Age Group, $F(1, 139) = 14.73$, $p < .001$, $\eta_p^2 = .096$, such that younger children indicated liking items more than older children. This is likely due to the fact that items included on the POAT-PM target younger children, as this measure was developed for preschoolers. There was also a significant main effect of Child's Gender $F(1, 139) = 4.09$, $p = .045$, $\eta_p^2 = .029$, and a significant Personal Measure X Gender interaction, $F(1, 139) = 166.48$, $p < .001$, $\eta_p^2 = .55$, which showed that boys preferred masculine items and girls preferred feminine items (Figure 12, Figure 13). In fact there was a significant moderate correlation between preference for masculine and feminine items, $r = -.324$, $p < .001$, suggesting that children were less likely to prefer masculine activities if they preferred feminine activities, and vice versa.

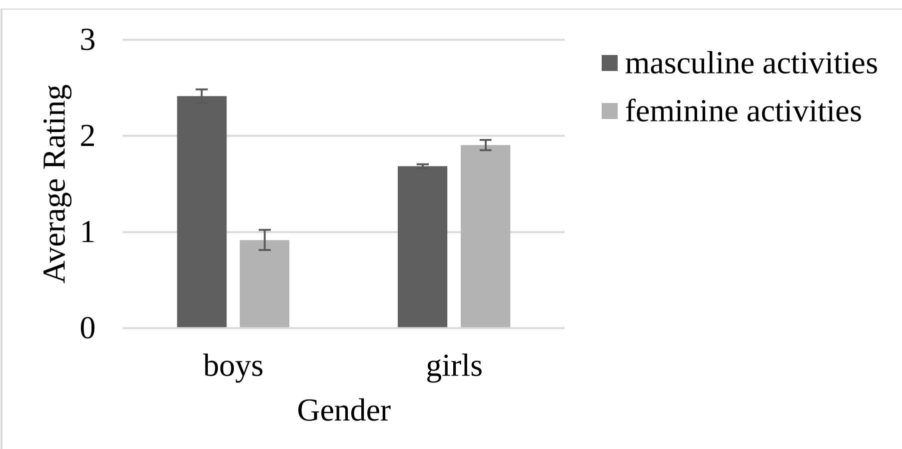

Figure 12. POAT-PM by Gender. Children's preferences for masculine and feminine activities by gender. Ratings are on a three-point Likert scale (1=I don't like it much, 2=I like it some, 3=I like it a lot). Error bars denote standard error of the mean.

When children retold the story of the block-competition winner, 69% of children used a male pronoun to refer to the child in the story, $\chi^2(1, N = 154) = 99.9$, $p < .001$ (four children did not answer this question). Older children were more likely to use a male pronoun and boys were more likely than girls to use a male pronoun, χ^2s ≥ 8.3, *p*s ≤ .004 (Figure 14). These results suggest that children implicitly believe that boys are better at building with blocks than girls.

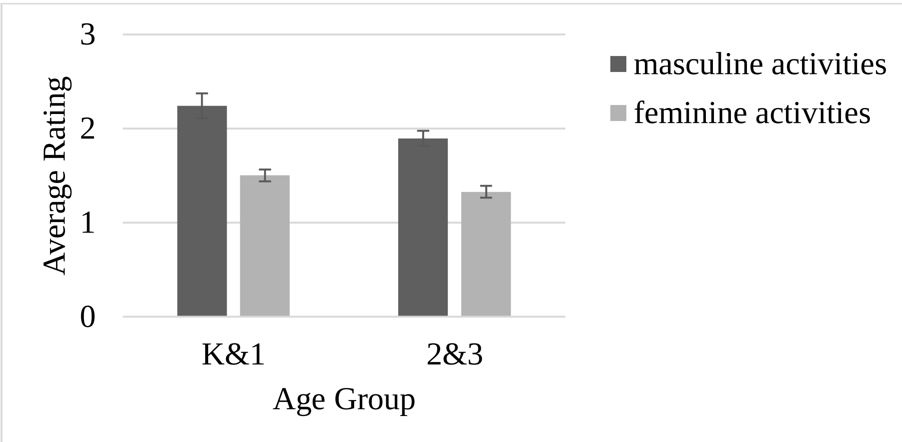

Figure 13. POAT-PM by Age Group. Children’s preferences for masculine and feminine activities by age group. Ratings are on a three-point Likert scale (1=I don’t like it much, 2=I like it some, 3=I like it a lot). Error bars denote standard error of the mean.

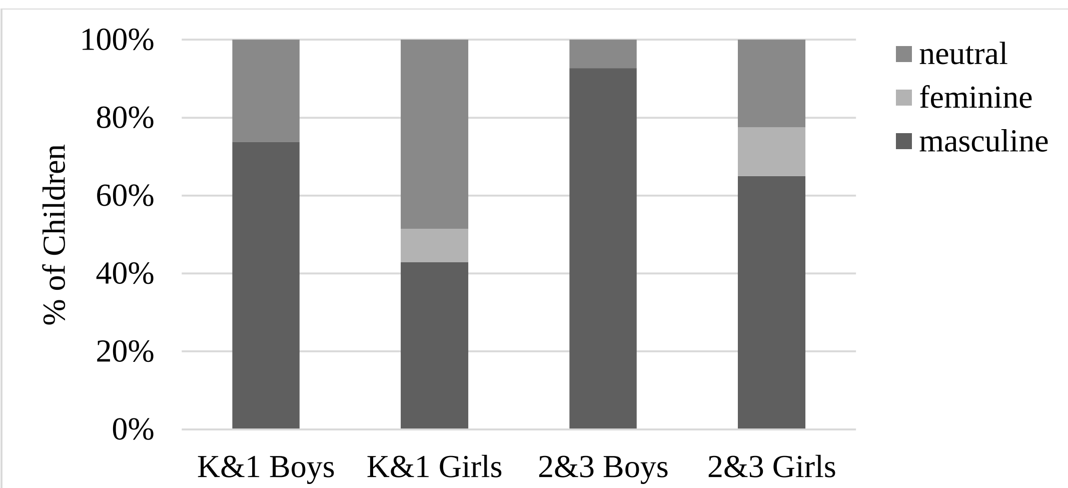

Figure 14. Experiment 2: Implicit Stereotypes. Percentage of children using masculine, feminine, or neutral pronouns to refer to the block-contest winner.

Looking at children’s explicit stereotypes about blocks, overall, children had stereotypical views about all three types of blocks (χ^2s ≥ 27.1, *p*s < .001); they were more likely to respond with “boys only” for multicolored and blue blocks and with “girls only” for pink blocks. Whereas boys’ and girls’ responses for pink blocks were similar ($\chi^2(1, N = 158) = 0.93$, $p = .3$), boys were more stereotypical than girls (i.e., indicating “boys only”) when answering the question about multicolored blocks and blue blocks (χ^2s ≥ 6.54, *p*s $\leq .01$). Children were also more likely to cite “girls only” when indicating competence about building with pink blocks, suggesting that the color pink overturns the established stereotype of boys being better at building with LEGO blocks (Figure 15). Analysis of children’s responses to the explicit block stereotype questions by age group found that older children were more stereotypical when answering the question about multicolored blocks ($\chi^2(1, N = 158) = .15$, $p = .042$), but their response patterns were similar for blue and pink blocks (χ^2s ≤ 0.59, *p*s $\geq .44$).

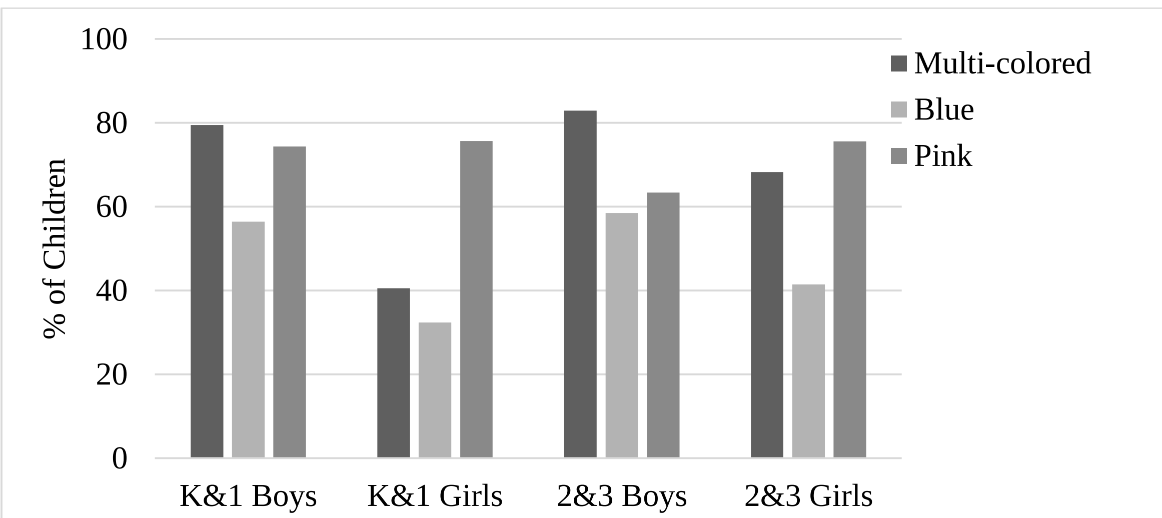

Figure 15. Experiment 2: Explicit Block Stereotypes. Percentage of children who gave a stereotypical answer for multi-colored (“boys only”), blue (“boys only”) and pink blocks (“girls only”).

Parent Questionnaire

Among parents who indicated their child’s favorite toys ($n = 144$), 28% of boys’ parents

(20 boys) indicated LEGO blocks as their favorite toy whereas parents of only one girl mentioned LEGO blocks as an additional favorite toy. To further analyze this data, these answers were coded into seven different categories based on the type of toys indicated (Table 2). A second coder, blind to the hypotheses and children's gender also coded these responses. Inter-coder reliability was very high (κ = .92). When parents indicated more than one toy, only the first toy mentioned was analyzed in this section. The most frequently cited categories among boys were action figures; weapons and vehicles; and blocks and construction. The least cited categories among boys were arts and crafts and dolls and stuffed animals. Parents of girls were most likely to cite dolls and stuffed animals and arts and crafts, and least likely to cite sports, action figures, and blocks. This pattern follows stereotypical toy preferences presented by Goble et al. (2012).

	Boys (n = 73)	Girls (n = 71)
Electronics	16.4	7.0
Dolls/stuffed animals	4.1	59.2
Action figures/vehicles/weapons	35.6	2.8
Arts and crafts	2.7	21.1
Blocks/construction	30.1	0
Sports	11.0	1.4
other	0.0	8.5

Table 2. Experiment 2: Favorite Toys. Boys' and girls' favorite toys as indicated by parents coded into seven different categories. Values represent percentage of children.

Looking at the availability of LEGO blocks, dolls, cars and trucks, and kitchen and tea sets (Table 3), 51% of boys' parents and 71% of girls' parents reported that all four types of toys were available in their homes. According to parents, their child's frequency of play with these toys followed stereotypical patterns, where parents indicated that boys were more likely to play with LEGO blocks and cars/trucks, and girls were more likely to play with dolls and kitchen/tea sets (all *t*s ≥ 3.38, all *p*s ≤ .001). Parents may have been conforming to gender norms rather than reporting actual frequencies, especially given that meta-analytic research has shown that parents consistently encourage gender-typed activities for their children (Lytton & Romney, 1991). However, even if this were the case, these results are telling of what parents in this population expect of their children and how this may affect children's play preferences and later preferences.

	Boys		Girls	
	Availability	Frequency	Availability	Frequency
LEGOs	97%	2.37	84%	1.79
Dolls	71%	1.18	100%	2.52
Cars/Trucks	99%	2.21	84%	1.51
Kitchen/Tea Set	56%	1.47	93%	2.11

Table 3. Availability and Frequency of Play with Toys. Parent questionnaire: percentage of boys and girls who have LEGO blocks, dolls, cars/trucks, and kitchen/tea sets in their homes and how frequently they play with them. Frequency is based on a five-point Likert scale, where 1 = never and 5 = all the time.

Block Construction, Stereotyping, and Toy Preferences

A multiple regression examined whether children's speed on the LEGO task could be predicted from stereotyping level on POAT-AM, implicit and explicit block stereotype, the child's favorite toy as indicated by parents (LEGO or not LEGO), and how often children played with LEGO blocks, after controlling for child's gender and age group. The model including only Gender and Age Group significantly predicted 21% of the variance (R^2 = .207, *F*(2, 127) = 16.34, *p* < .001). Adding the remaining five variables mentioned above resulted in a total prediction of 37% of the variance (R^2 = .369, *F*(2, 127) = 10.02, *p* < .001). In the first model, Gender significantly predicted speed (β = .22, *p* = .007), as did Age Group (β = -.39, *p* < .001). In the second model, children's implicit block stereotype and stereotyping level on POAT-AM did not significantly predict speed (*p*s ≥ .53). However, explicit block stereotype (β = .29, *p* = .001), child's favorite toy (β = -.23, *p* = .007), and how often children played with LEGO blocks as indicated by parents (β = -.19, *p* = .017) were significant predictors. These results suggest that after controlling for the effects of age and gender, both boys and girls were slower on the block-construction task if they thought boys were better at building with LEGO blocks, if LEGO blocks were not their favorite toy, and if they did not play with LEGO blocks frequently.

Taken together, the results from Experiment 2 suggest several important findings. First, there was no strong evidence of stereotype threat for girls in this sample of elementary-school girls. In light of having found stereotype threat in a younger sample with a similar task (Experiment 1), there are a number of possible explanations for this result. First, this sample was taken from different schools in a different Midwestern city that could potentially have reacted differently to gender-identity activation. However, the stereotype assessment results suggest that children in this sample were in fact more stereotypical than girls in Experiment 1, which should have increase girls' vulnerability to stereotype threat (Ambady et al., 2001, Galdi et al., in press). Thus, it is unlikely that the different location explains the difference in results. Another possible explanation is that preschoolers are susceptible to stereotype threat but elementary-school children are not. This explanation seems unlikely, given that other studies (e.g., Ambady et al., 2001; Galdi et al., in press; Neuville & Croizet, 2007) have shown stereotype threat in older children. Also, as children learn stereotypes, these stereotypes are more likely to have an effect on their performance (Galdi et al., in press). A third explanation is that the task was too simple for elementary-school children. Some studies suggest that stereotype threat acts differently on easy and difficult tasks (Neuville & Croizet, 2007): girls sometimes show stereotype threat on difficult tasks but not on easy ones. A fourth explanation is that stereotype threat is already manifested in these results, as evident by the discrepant performance of boys and girls, and that the gender activation manipulations did not influence girls' performance beyond the threat that was already present. Merely mentioning that this is a hard game involving blocks (an activity that elementary-school children already associate with being masculine) may have created a threat for all the girls in the study, thus affecting their performance. This finding and interpretation is consistent with Ganley et al.'s (2013) work with children in fourth through twelfth grade. In their stereotype threat study, girls consistently underperformed compared to boys and they suggested they may be showing continuous stereotype threat regardless of condition. This explanation, together with the proposition that stereotype threat is not manifested on easy tasks, is supported by the narrowing gap between performance by boys and girls on the block-construction task as they get older. It may be the case that younger girls felt more threatened as the task was harder for them and that older girls found it easier and thus were able to overcome stereotype threat more easily (although gender differences were still evident). These results will be further discussed in Chapter 4.

CHAPTER 3

STUDY 2: CHILDREN'S ATTITUDES ABOUT COMPETENCE AND LEARNING ABILITY

Although boys and girls start out performing similarly on STEM-related school subjects, especially math, by high school, boys outperform girls (Hyde & Linn, 2006). The gap widens further at college level and in the job market (NSF, 2013). One explanation for this phenomenon is that women's stereotypical beliefs about their abilities create a self-stereotype threat that hinders their performance. Girls and women may also be threatened by others' (men's and women's) stereotypical beliefs and expectations. As is evident from Study 1, preschoolers' performance is susceptible to stereotype threat and girls' performance on a simple task for their age is affected by their beliefs about boys' and girls' abilities. Therefore, Study 2 further explores boys' and girls' beliefs about males' and females' competence and learning ability on STEM and non-STEM fields. Boys and girls who participated in Experiment 2 of Study 1 completed a series of tasks where they rated the competence of men and women and the learning ability of boys and girls on STEM and non-STEM fields. They also rated the difficulty of STEM and non-STEM professions and indicated their school-subject preferences.

In light of the literature reviewed in Chapter 1, I hypothesized that both boys and girls would rate males as more competent at and more adept at learning in STEM-fields than females and that children's ratings would be similar for males and females on non-STEM fields, mirroring the representation of men and women in STEM and non-STEM fields in real life. I expected girls to have lower stereotypical ratings than boys, given that in-group bias competes with STEM stereotypes (see Chapter 1 for more detailed discussion of this hypothesis). I expected children in the older age group to have more stereotypical STEM attitudes, as children would have had more experience with different professions and would be more knowledgeable about related stereotypes (Ruble et al., 2006). Based on the research literature discussed in Chapter 1 (e.g., Cvencek et al, 2011; Entwistle et al., 1987), I expected to see gender differences in children's preferences for school subjects and their career aspirations.

Method

Participants

A subset of children who participated in Experiment 2 (N = 132, 64 girls) participated in this study (it was not possible to conduct a second session with children from the private school). This data was collected in a separate session, which occurred one to five weeks after participation in Experiment 2. Children were divided into the same two age groups as in Experiment 2: Children in kindergarten and first grade (n = 58, M = 6.34, SD = 0.58, range = 5.19 - 7.26, 28 females) and children in second and third grade (n = 74, M = 8.29, SD = 0.64, range = 7.18 - 10.00, 36 females).

Procedure and Materials

Children were tested in groups of two to five, and most groups had mixed genders. Children were seated such that they could not see what other children were writing. The experimenter explained the instructions and read the questions and children were provided with a booklet where they indicated their answers. Sessions lasted between fifteen and twenty minutes. The study consisted of four main sections in the following order: Competence ratings, difficulty ratings, learning ability ratings, and school-subject preferences.

Competence Ratings: During the first part of the study, the experimenter explained that adults work in different jobs and that some people are better at their jobs than others (see Appendix G for full script). Then, the experimenter trained children to use a six-point scale (using stars) to rate how good adults are at their jobs. For each adult introduced, children had to fill in the number of stars they thought the adult deserved. Five stars meant the person was excellent at the job and one star meant the person was not good at all at the job. Children were also given the option to write a zero if they thought that the person absolutely did not deserve any stars. After explaining what each number of stars represented, the experimenter asked children to decide how many stars a person deserved if s/he was excellent or not good at all at the job to make sure that children understood the scale. Then, children proceeded through the booklet and rated how good a man and a woman were at each profession. There was a total of eight professions, four STEM (biologist, computer scientist, engineer, and mathematician) and four non-STEM (journalist, historian, sociologist, and advertiser). The choice of STEM professions followed closely from the STEM abbreviation. Non-STEM professions were chosen from a list of majors offered at a large university. The professions had to be easily defined for and understood by elementary-school students. Each profession was presented on a separate page of the booklet and was represented by a picture of objects involved in the profession (e.g., a computer for the computer scientist and a variety of animals for the biologist). Underneath the picture, there was a male and a female stick figure (see Appendix H for sample pages) above a blank set of five stars. For every profession, the experimenter explained what the profession entails (e.g., "This man and this woman are engineers. An engineer knows about machines. An engineer can design and build new machines and can fix machines and make them work better."). She then asked children to decide how good they thought this man and this woman were at their job by filling in the number of stars they thought the man and the woman deserved. Professions were listed such that STEM professions alternated with non-STEM professions. There were two orders of professions: one started with a STEM profession and the second with a non-STEM profession. On half of the forms, the man was on the left and the woman was on the right and vice versa. Thus, there were four different versions of the booklet. Children in each session all had the same version. At the end of this segment, children were asked to write down what they would like to be when they grow up and were instructed to think of anything they want and not restrict themselves to the professions discussed in this segment. Children were instructed to request help if they were unable to spell words.

Difficulty Ratings: To gauge whether children perceived STEM and non-STEM professions differently, children rated how difficult they thought each of the professions was for adults ("how hard do you think it is to be an engineer for grownups?") on a five-point scale by circling a number from 1-5, one being very easy and five being very hard. Children were introduced to the scale and were trained to use it via example questions (e.g., "if you think a job is very easy, which number would you circle?"), then children had to circle a number for every profession. Each profession was featured on a separate page, and each page included the same picture symbolizing the profession from the competence-ratings section. The order of professions corresponded to the order in the previous section. For every profession, the experimenter stated the name of the profession and children were instructed to inquire if they did not remember what the profession was about.

Learning-Ability Ratings: After taking a brief break, children were told that some fifth-grade boys and girls took classes to learn about new, difficult topics and then took a difficult five-question test. Children were asked to indicate how many questions they thought each boy and girl got correct on the test on a six-point scale by circling a number similar to the difficulty-ratings task. In this case, the numbers 1-5 represented the number of questions a boy or a girl answered correctly on the test. Children were also given the option to write a zero if they thought the boy or girl did not answer any questions correctly. After children were trained to use the scale, they indicated their learning-ability ratings for eight different subjects, four STEM (physical sciences, life sciences, computers, math) and four non-STEM (grammar, social studies, music, art). These subjects and their descriptions were adapted from the fifth-grade curriculum of the school district where data was collected. On each page, a different subject was featured with a picture representing the topic and a male and female stick figure below it with the numbers 1-5 under each stick figure (see Appendix H for sample pages). The experimenter introduced each topic by labeling the subject and explaining what the course was about (e.g., "this boy and this girl took a hard class to learn all about advanced math. They learned all about algebra and geometry including fractions and learned how to calculate the area and surface of different shapes.") Similar to the competence-ratings task, subjects were presented in one of two orders and STEM and non-STEM fields were alternated.

School-Subject Preferences: In the last portion of the study, children were introduced to five faces and were trained to use them to describe how much they like things they learned at school ("if you like something a lot, which face would you circle?"). Then, they had to circle the face that best described how much they liked each of seven school subjects (science, math, reading, spelling, art, music, and computers) which appeared on one page in a list with faces next to each subject. Finally children had to circle their favorite school subject from a list on the last page of the booklet.

Results and Discussion

Competence Ratings

To analyze competence ratings, a composite score was created for each of the following four categories by adding children's ratings and dividing by the number of items per category (4): females in STEM professions (FemaleSTEM, $\alpha = .62$), females in non-STEM professions (FemaleNonSTEM, $\alpha = .64$), males in STEM professions (MaleSTEM, $\alpha = .48$), and males in non-STEM professions (MaleNonSTEM, $\alpha = .50$). Means and standard deviations for all variables of Study 2 are presented in Table 4 (see Table 9 in Appendix I for means of individual items).

	M	SD
Competence Ratings		
FemaleSTEM	3.53	1.08
FemaleNonSTEM	3.49	1.12
MaleSTEM	4.12	0.84
MaleNonSTEM	3.62	0.95
Difficulty Ratings		
DifficultySTEM	3.33	0.98
DifficultyNonSTEM	3.14	0.88
Learning-Ability Ratings		
GirlSTEM	3.16	1.12
BoySTEM	3.52	1.00
GirlNonSTEM	3.64	1.08
BoyNonSTEM	3.63	0.95
LikeSTEM	3.01	0.79
LikeNonSTEM	2.98	0.83

Table 4. Competence, Difficulty, and Learning Ratings. Means and standard deviations of competence, difficulty, and learning ability ratings as well as children's personal preferences for STEM and non-STEM school subjects. All scores are out of five points (the range for competence and learning ability was 0-5, whereas the scale for difficulty and for subject preference was 1-5).

A repeated-measures ANOVA was conducted with Competence Ratings (FemaleSTEM, FemaleNonSTEM, MaleSTEM, MaleNonSTEM) as the within-subjects variable and Child's

Gender (male, female) and Age Group (K&1, 2&3) as the between-subjects variables. Mauchly's test indicated that the assumption of sphericity was violated, $\chi^2(5) = 16.33$, $p = .006$, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity, $\varepsilon = .97$ (Huynh & Feldt, 1976). The ANOVA revealed a significant main effect of Competence Ratings, $F(2.91, 128) = 16.13$, $p < .001$, $\eta_p^2 = .11$. Pairwise comparisons revealed that children's ratings for males in STEM professions were significantly higher than ratings for the remaining three categories (*p*s < .001).

There were no significant main effects of Child's Gender, $F(1, 128) = 1.30$, $p = .26$, $\eta_p^2 = .01$ or Age Group $F(1, 128) = 3.03$, $p = .08$, $\eta_p^2 = .02$. However, there was a Competence Rating X Child's Gender interaction, $F(2.91, 128) = 21.83$, $p < .001$, $\eta_p^2 = .15$. Girls' and boys' ratings were significantly different from each other on all categories (*t*s > 2.68, *p*s < .008) except on males in non-STEM professions ($p = .14$). In each category, boys' ratings were higher for males and girls' ratings were higher for females (Figure 16). However, planned paired-samples *t*-tests looking at boys' and girls' ratings of males and females in STEM and non-STEM professions showed that boys rated males significantly higher than females on both STEM and non-STEM professions (*t*s > -4.78, *p*s < .001) whereas girls' ratings were significantly higher for females than males on non-STEM professions ($t(63) = 3.61$, $p = .001$), but they were not significantly different on STEM professions ($t(63) = -0.72$, $p = .48$). Thus, boys always rated males as more competent, but girls rated females as more competent only on non-STEM professions.

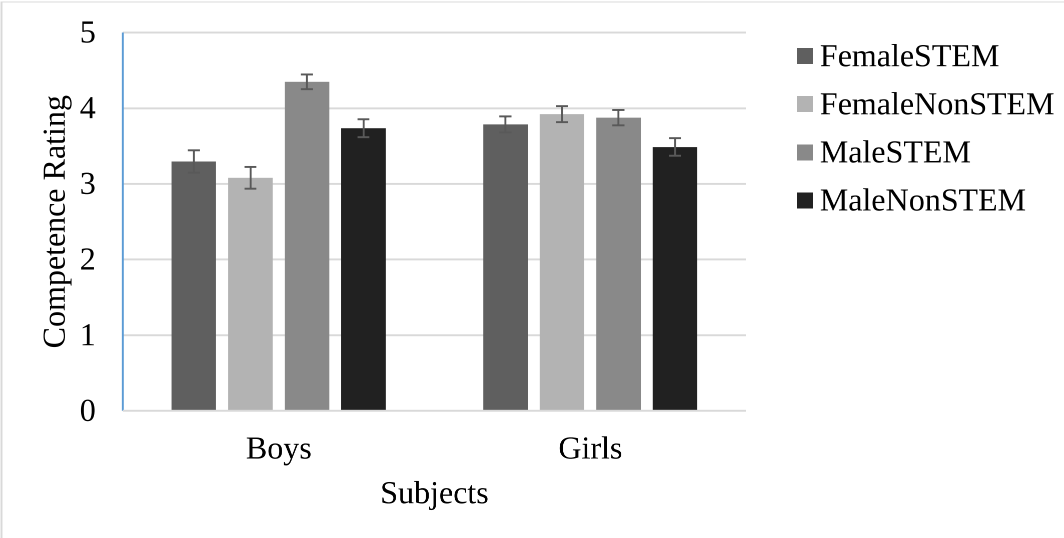

Figure 16. Competence Ratings by Gender. Boys' and girls' competence ratings for males and females in STEM and non-STEM professions. Error bars denote standard error of the mean.

There was also a Competence Rating X Age Group interaction, $F(2.91, 128) = 3.16$, $p = .026$, $\eta_p^2 = .02$ (Figure 17). Children in the younger and older age group had similar competence ratings for all categories (*t*s < 1.07, *p*s > .29) except for rating males in STEM professions ($t(103.9) = -3.82$, $p < .001$), such that second and third graders rated males in STEM professions as significantly better at their profession than younger children.

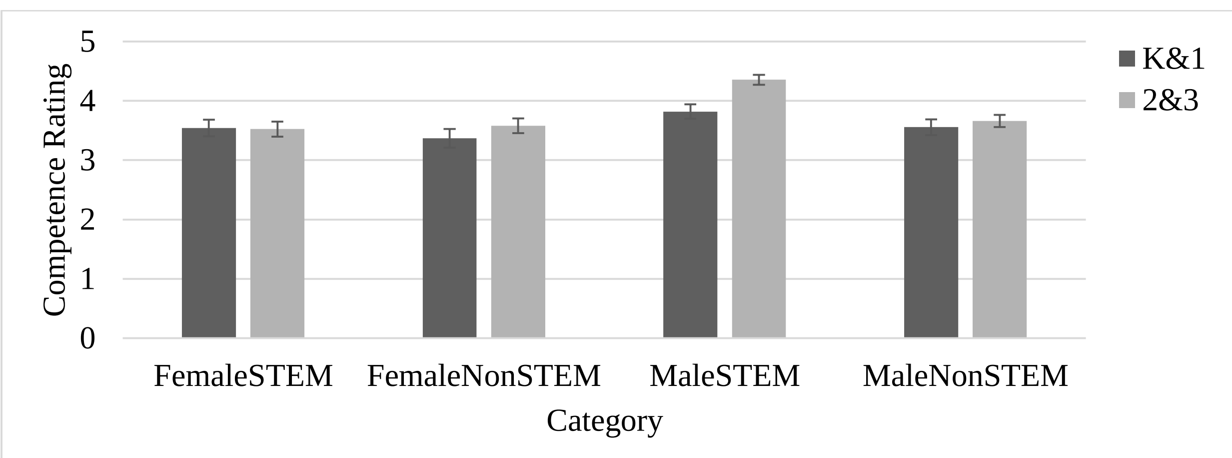

Figure 17. Competence Ratings by Age Group. Competence ratings on the four different categories by age group. Error bars denote standard error of the mean.

To test whether there were gender differences across age groups, planned pairwise comparisons revealed that among the younger age group, children's ratings showed an in-group bias such that boys' ratings were higher for their own gender than for the opposite gender regardless of category (*t*s > -3.37, *p*s < .002). The same trend was apparent for girls' ratings in the younger age group, although their ratings were not significantly different between males and females in STEM fields (STEM: $t(27) = 1.51$, $p = .14$; non-STEM $t(27) = 2.47$, $p = .02$). On the other hand, for the older age group, children showed an in-group bias for non-STEM professions (*t*s > 2.65, *p*s < .012), but both boys and girls rated males significantly higher on STEM professions than females (*t*s > -2.38, *p*s < .023). Thus, whereas boys' ratings across age groups were consistently higher for males than for females, girls seem to have developed stronger stereotypes about STEM professions in the older age group, such that they rated men and women similarly on STEM professions in the younger age group, but their ratings were higher for males than females on STEM professions in the older age group.

Taken together, results from children's competence ratings support my hypotheses that children assign higher competence ratings to males in STEM professions and suggest that there are two main forces that determine children's perceptions of individuals in STEM and non-STEM fields: in-group bias and gender stereotypes. Younger children showed an in-group bias for their own gender and believed that members of their own gender were better at all professions than members of the opposite gender (although girls' ratings were not significantly different for males and females in STEM professions, suggesting a competition between in-group bias and stereotypes). Older children continued to show in-group bias on non-STEM professions, with both boys and girls indicating that members of their own gender would be more competent at these jobs. Boys' ratings were also higher for males than females. However, boys' gender matches the STEM stereotype, thus it is unclear whether their answers were motivated by in-group bias only or whether gender stereotypes were also affecting their decisions. Second- and third-grade girls' answers, on the other hand, showed a stark change from younger girls' answers. The older girls indicated that males are better at STEM professions than females, thus abandoning their in-group bias in favor of learned gender-stereotypes about STEM fields.

Difficulty Ratings

To analyze difficulty ratings, children's answers (on a 1-5 scale) were averaged to create two variables: difficulty of STEM professions and difficulty of non-STEM professions (difficulty ratings of individual professions are provided in Table 5). A repeated-measures ANOVA with Difficulty Rating (STEM and non-STEM) as the within-subjects variable and Child's Gender and Age Group as the between-subjects variables revealed no significant main effects of Difficulty Rating, Child's Gender, or Age Group (Fs ≤ 3.30, ps $\geq .072$). However, there was a significant Difficulty Rating X Age Group interaction, $F(1, 125) = 9.38$, $p = .003$, $\eta_p^2 = .07$ (Figure 18). children in the younger age group rated STEM and non-STEM professions similarly ($t(57) = -0.73$, $p = .47$), whereas children in the older age group rated STEM professions as significantly more difficult than non-STEM professions ($t(70) = 4.40$, $p < .001$). In light of children's competence ratings (discussed above), the difficulty ratings suggest that as children grow older they believe that STEM professions are more difficult and at the same time the gap between ratings for males and females in STEM professions widens. However, there was no correlation between difficulty ratings of STEM professions and ratings of males or females in STEM professions (rs $\leq .04$, ps $\geq .55$), possibly suggesting that on an individual level, children were not considering the difficulty of each profession as they were rating the competence of males and females.

STEM Professions	M	SD	Non-STEM Professions	M	SD
Computer Scientist	3.13	1.48	Advertiser	2.68	1.48
Biologist	3.19	1.48	Journalist	2.83	1.42
Mathematician	3.42	1.62	Sociologist	3.37	1.41
Engineer	3.54	1.50	Historian	3.62	1.41
Overall STEM	3.33	0.98	Overall Non-STEM	3.14	0.88

Table 5. Difficulty Ratings: Individual Professions. Children's Difficulty ratings of STEM and non-STEM professions sorted from easiest to hardest.

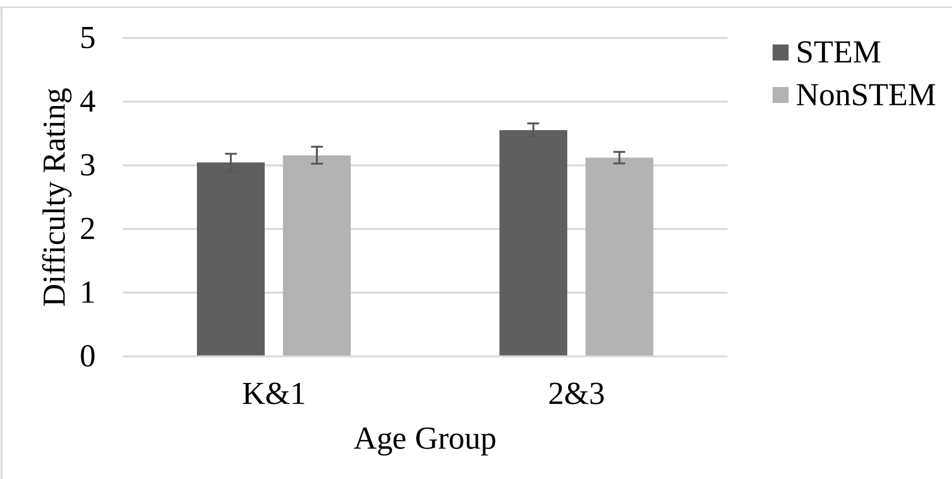

Figure 18. Difficulty Ratings by Age Group. Children's ratings of the difficulty of STEM and non-STEM professions by age group. Error bars denote standard error of the mean.

Learning-Ability Ratings

Children's learning ability ratings were scored similar to competence ratings, such that four categories were created representing children's ratings of boys' and girls' number of questions answered correctly on STEM and non-STEM subjects (GirlSTEM [α = .68], GirlNonSTEM [α = .69], BoySTEM [α = .59], BoyNonSTEM [α = .57]). A repeated-measures ANOVA was conducted with Learning Ability Rating (GirlSTEM, GirlNonSTEM, BoySTEM, BoyNonSTEM) as the within-subjects variable and Child's Gender (male, female) and Age

Group (K&1, 2&3) as the between-subjects variables. Mauchly's test indicated that the assumption of sphericity has been violated, $\chi^2(5) = 35.10$, $p < .001$, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity, $\varepsilon = .90$ (Huynh & Feldt, 1976). The ANOVA revealed a significant main effect of Learning Ability, $F(2.70, 122) = 16.13$, $p < .001$, $\eta_p^2 = .06$. Pairwise comparisons revealed that children's ratings were significantly lower for girls learning about STEM topics than all other categories (*p*s < .004, Figure 19).

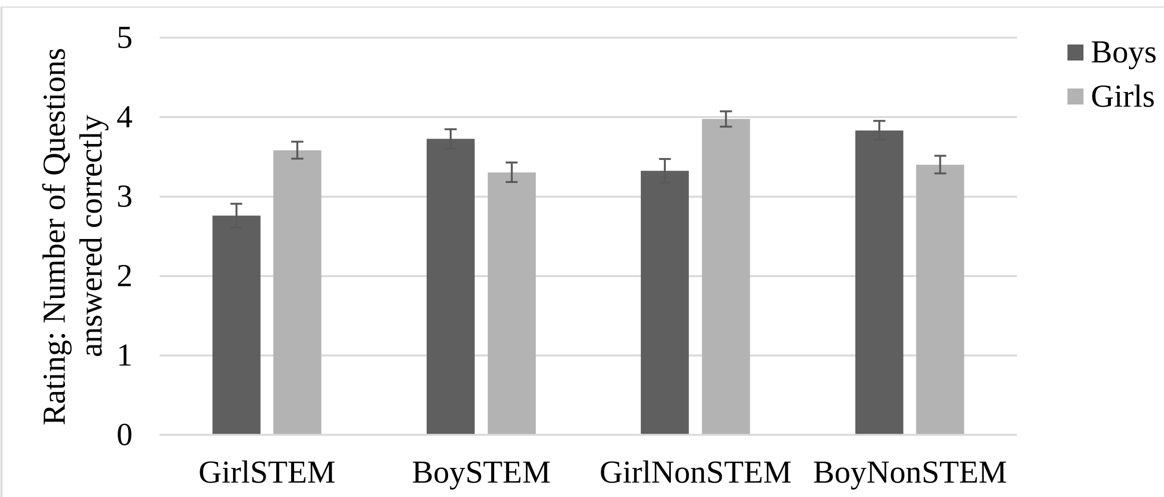

Figure 19. Learning Ability Ratings by Gender. Boys' and girls' ratings of boys and girls learning about new STEM and non-STEM subjects. Error bars denote standard error of the mean.

There were no significant main effects of Gender, $F(1, 122) = 1.49$, $p = .23$, $\eta_p^2 = .01$ or Age Group $F(1, 122) = 0.09$, $p = .77$, $\eta_p^2 = .001$. However, there was a Learning Ability Category X Child's Gender interaction, $F(2.70, 122) = 19.87$, $p < .001$, $\eta_p^2 = .14$. Girls' and boys' ratings were significantly different from each other on all categories (*t*s > 2.41, *p*s < .017), such that in each category, boys' ratings were higher for males and girls' ratings were higher for females. Planned paired-samples *t*-tests looking at boys' and girls' ratings of boys and girls in

STEM and non-STEM subjects (Bonferroni-adjusted *p*-value to achieve significance for these four *t*-tests is $p = .0125$) showed that boys rated males significantly higher than females on both STEM and non-STEM subjects (*t*s > -2.94, *p*s < .005) whereas girls' ratings were significantly higher for females than males on non-STEM subjects ($t(61) = 5.37$, $p < .001$), but not on STEM subjects ($t(61)= 2.05$, $p = .045$). Additionally, comparing children's ratings for girls learning about STEM and non-STEM subjects revealed that both boys and girls thought that girls would answer more questions correctly when tested on learning of non-STEM topics than on STEM topics (*t*s > 3.28, *p*s < .002).

Thus, children's learning ability ratings suggest that they are strongly affected by in-group bias, where boys rate males higher than females and girls rate females higher than males (on non-STEM only). However, overall children's ratings were lower for girls learning STEM subjects than the other three categories. There was also no developmental differences between both age groups (see Table 10 in Appendix I for means of individual items).

Preferences and Career Aspirations

Children's preferences for school subjects were aggregated into STEM and non-STEM composites. A repeated-measures ANOVA was conducted to analyze children's preferences for STEM (math, science, and computers) and non-STEM (reading, spelling, art, and music) school subjects, with Category (STEM, non-STEM) as the within-subjects variable and Child's Gender and Age Group as between-subjects variables. The ANOVA revealed no main effect of Category, Gender, or Age Group (*F*s < 3.52, *p*s > .063). However, there was a significant Category X Child's Gender interaction $F(1, 127) = 12.36$, $p = .001$, $\eta_p^2 = .09$, such that boys' and girls' preference ratings for STEM subjects were similar ($t(129) = -0.47$, $p = .64$) but girls' ratings for non-STEM subjects were significantly higher than boys' ratings ($t(129) = -3.53$, $p = .001$). This effect is likely driven by girls' stronger preference for art and music, as a composite for non-STEM including reading and spelling only did not yield a significant difference between boys' and girls' ratings ($t(130) = -1.67$, $p = .10$, Figure 20).

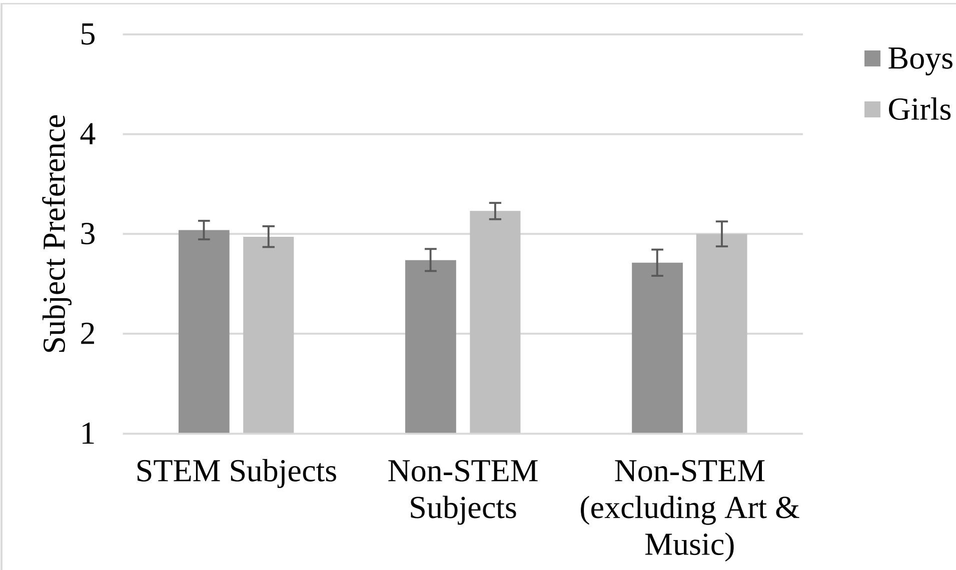

Figure 20. School-Subject Preferences by Gender. Boys' and girls' preference ratings for STEM and non-STEM school subjects. Values represent Likert scale where 1 = don't like at all and 5 = like it a lot. Error bars denote standard error of the mean.

When children were asked to circle their favorite school subject, boys' most cited subject was computers (27%) followed by math (21%) and the least cited were spelling and music (3% each). Girls' most cited subject was art (43%), followed by science (21%) and the least cited was spelling (3%). Thus, boys' two favorite subjects were STEM subjects, whereas girls' favorite subject was a non-STEM subject followed by science.

To analyze boys' and girls' career aspirations, children's answers were coded twice: once coding whether or not their answer was a STEM field and the second coding was for different profession categories. Both coding schemes were also implemented by a second coder unaware of the study's hypotheses or children's gender and there was high inter-rater reliability ($\kappa = .81$

and κ = .92, respectively). Discrepancies were resolved through discussion. Chi-square tests revealed no age or gender differences on the proportion of children indicating a STEM-related career aspiration (χ^2s ≤ 1.3, *p*s ≥ .25). Overall, 24% of children aspired to have a STEM-related career. The most frequently cited STEM professions were scientist and veterinarian and the most frequently cited non-STEM professions were teacher and police officer. (see Table 11 in Appendix J for full list of career aspirations).

	Boys	Girls
Athletics	29.0	10.9
Arts	4.3	10.9
STEM	24.6	17.2
Law enforcement	21.7	9.4
Agriculture/Animal Care	13.0	12.5
Education & Services	7.2	35.9
Homemaking	0.0	3.1

Table 6. Career Aspiration Categories. Percentage of boys and girls indicating career aspirations in each of seven different categories.

In addition to examining whether or not children's career aspirations were STEM-related, I analyzed whether they reflected gender stereotypes. To do so, children's answers were coded into seven different categories: athletics (e.g., professional athlete, basketball player), arts (e.g., fashion designer, artist), STEM (e.g., doctor, engineer), law enforcement (e.g., policeman, FBI agent), agriculture and animal care (e.g., veterinarian, farmer), education and services (e.g., teacher, sales person), and homemaking (e.g., mom). Table 6 shows children's answers coded into these categories. Boys were more likely to express interest in professions related to sports, STEM, and law enforcement, whereas girls were more likely to prefer professions in education and services. In fact, boys' and girls' response patterns were significantly different (χ^2 (6, *N* =

133) = 26.48, $p < .001$). There were no significant differences in response patterns by age group (χ^2 (6, N = 133) = 11.68, $p = .07$). Interestingly, both boys and girls showed similar levels of interest in agriculture and animal care, possibly because the sample originated from a rural area.

Comparing Attitudes and Preferences

In order to assess whether children's preferences are consistent with their attitudes, correlations were calculated between children's school-subject preferences and their ratings of girls and boys learning about school subjects. The results suggest that for both boys and girls, higher preference for STEM school subjects was moderately correlated with higher ratings for girls learning about STEM subjects (boys: $r = .31$, $p = .012$; girls: $r = .36$, $p = .003$). However, children's STEM preferences did not correlate with ratings for boys learning about STEM subjects (rs ≤ .11, ps ≥ .37). Looking at the relationship between children's preferences for non-STEM subjects and rating of boys' and girls' learning about non-STEM subjects, boys' non-STEM preferences correlated moderately with higher ratings for both girls learning about non-STEM subjects ($r = .36$, $p = .003$) and boys learning about non-STEM subjects ($r = .31$, $p = .01$), but girls' non-STEM preferences did not correlate with neither (rs ≤ .06, ps ≥ .60). Thus, it seems likely that children who liked STEM subjects more were also more likely to expect fifth-grade girls to do well on tests of similar nature, whereas boys who liked non-STEM subjects were more likely to expect fifth-grade boys and girls to do well on non-STEM tests. As this is correlational data, causality can only be suggested and these results could be interpreted two ways: it could be that children who like STEM subjects have more egalitarian views, thus assigning higher ratings to girls as well. Alternatively, it may be possible that as children like school subjects more, they feel these subjects are more accessible to other children as well, thus awarding higher grades to other children.

Stereotype Threat, STEM Attitudes, and Preferences

This section examines relationships between children's block-construction performance and stereotyping in Study 1 and children's attitudes towards and preferences for STEM and non-STEM fields in Study 2. Based on the literature presented in Chapter 1, I expected that children who like STEM subjects more, particularly math as it is related to block construction, would perform better on a block construction task. Contrary to this prediction, there was no significant relationship between speed on the block-construction task and boys' and girls' preference for STEM fields in general or liking math in particular (rs $\leq .04$, ps $\geq .65$).

I also predicted that children who had stereotypical attitudes about toys and activities would also have stereotypical attitudes about STEM fields. Comparing children's stereotypical attitudes on POAT-AM and their STEM attitudes in Study 2 revealed a significant negative correlation between level of stereotyping on POAT-AM and scores awarded to girls learning about STEM ($r = -.17$, $p = .05$), but there was no relationship between POAT-AM stereotyping level and ratings awarded to females in STEM professions. Thus, children who had higher stereotypical attitudes on the POAT-AM were also more likely to award fewer points to girls learning STEM subjects, but their stereotypical attitudes did not predict their ratings of the competence of adult women in STEM professions. This suggests that children may be distinguishing between children and adults by showing similar trends for children but not adults (this possibility will be further discussed in Chapter 4).

There was also a significant correlation between children's preferences for masculine activities (POAT-PM) in Study 1 and their preference for STEM school subjects in Study 2, r = .21, p = .017. To eliminate the possibility that this relationship was due to a third variable - gender - correlating with both variables, this relationship was analyzed for boys and girls separately. The analysis revealed no significant relationship between preference for masculine activities and STEM school subjects for boys, r = .09, p = .49. Interestingly, there was a significant moderate correlation between girls' preference for masculine activities and their preference for STEM school subjects, r = .30, p =.017. Thus, girls who prefer stereotypically masculine play activities also show a preference for STEM subjects.

Taken together, Study 2 presents several key findings. First, overall, children attributed higher competence to males in STEM professions than females in STEM professions. This pattern was more pronounced in boys and in older children. Second, older children believed that STEM professions were more difficult for adults, although there was no consistent relationship between difficulty ratings and competence ratings. Third, children expressed more in-group bias when rating fifth graders' learning ability, resulting in boys mostly attributing higher learning ability to boys on most items and girls attributing higher ability to girls (although girls' ratings were not significantly different for males and females on STEM subjects). Boys' ratings suggest that they perceive a larger gap between boys' and girls' learning ability, especially on STEM fields, whereas girls' ratings suggest both in-group bias and STEM stereotypes are at play. With respect to preferences, boys and girls had similar school-subject preferences and similar likelihood of citing STEM career aspirations, although their career aspirations were stereotypical overall.

Comparing data from Study 1 and Study 2 also reveals two interesting findings. First, there seems to be a relationship between children's stereotyping level on the POAT-AM and their stereotypical rating of girls learning about STEM subjects where children who hold stronger stereotypical beliefs about toys and activities also do so for girls' STEM learning. Second, girls who show a stronger preference for masculine activities on the POAT-PM also show a stronger preference for STEM school subjects. Thus, there may be a relationship between the skills promoted by masculine toys and activities and STEM subjects. These results will be further discussed in Chapter 4.

CHAPTER 4

GENERAL DISCUSSION

The goal of this work was to shed light on potential precursors of the gender gap in STEM fields by studying the effects of gender stereotypes on children's attitudes and performance during the preschool and early elementary school years. In Study 1, children ages 4 through 9 were tested on a block-construction task under stereotype-threat or control conditions and their attitudes and preferences for masculine and feminine toys and activities were measured. Experiment 1 showed that, as hypothesized, preschool girls were susceptible to stereotype threat on a toy-based block-construction activity, even when they did not express explicit stereotypes about blocks. Experiment 2 suggested that contrary to the hypothesis, girls in kindergarten through third grade were not slower or less accurate in the stereotype-threat conditions compared to a control condition. However, there was an overall performance difference between boys and girls, where boys were significantly faster than girls. Further, children's performance on the block-construction task was predicted by their toy preferences and explicit stereotypes about blocks.

Study 2 examined a different potential precursor of women's underrepresentation in STEM fields by measuring children's attitudes towards females in STEM fields. Children rated how competent men and women were at STEM and non-STEM professions and how well fifth-grade boys and girls would do on a test of a recently learned STEM or non-STEM subject. As hypothesized, both boys and girls awarded lower competence ratings to women in STEM fields than to men and this discrepancy was larger in the older age group. Despite these stereotypical attitudes, both boys and girls expressed similar preferences for STEM school subjects and similar levels of STEM career aspirations. Comparing results for Study 1 and Study 2 also revealed that girls who had stronger preferences for masculine toys and activities were more likely to like STEM school subjects and that children who held stronger stereotypes about toys and activities were more likely to award fewer points to girls learning about STEM subjects. However, there was no evidence for a relationship between children's performance on the block-construction task and their STEM attitudes. In the following sections, I will critically discuss these findings in light of previous research and I will argue for their relevance to understanding the STEM gender gap.

Stereotype Threat: Emergence and Development

Preschool girls whose gender identity was activated worked more slowly on recreating shapes using LEGO blocks than girls whose gender identity was not activated (Exp. 1). This is the youngest age group studied thus far in the stereotype-threat literature, and showing that girls are susceptible to stereotype threat at this young age has important implications for understanding girls' trajectories. If girls are threatened by their own identity so early in life, they may avoid engaging in activities that make them uncomfortable (Frome et al., 2006). Because block construction is closely related to math (Mix & Cheng, 2012), this translates into girls potentially missing opportunities to improve their spatial skills. This then holds repercussions for the development of math skills (Cheng & Mix, 2014) and it may be one of the precursors of women's underrepresentation in STEM fields later on.

On the other hand, among older girls (Exp. 2), performance on the block-construction task when gender identity was activated did not differ from the control group. If stereotype threat is detectable as early as in preschool on a block-construction task, why is it not evident among children in kindergarten through third grade? There are several potential explanations for this inconsistency. It is possible that the task, which was originally developed for preschoolers and was slightly modified to be used with older children, was too easy for older girls and thus did not induce stereotype threat. This explanation is consistent with Neuville and Croizet's (2007) finding that girls' performance improved on easy math problems and was worse on harder problems. It could also be argued that the gender-identity activation task did not in fact activate the older girls' gender identity. However, this gender-activation task has previously been used to activate gender identity – and induce stereotype threat- in girls in kindergarten through third grade (Ambady et al., 2001; Neuville & Croizet, 2007). Thus, this explanation is unlikely. Another explanation is that older girls do not hold stereotypes about blocks and thus their gender-identity activation is irrelevant to the task. However, older girls expressed explicit and implicit stereotypical attitudes about blocks, and thus this explanation is also unlikely.

Alternatively, it is possible that girls were under continuous stereotype threat as a result of the experimenter describing the task as a hard task at the beginning of the experiment, even before the gender-identity activation. This description may have evoked anxiety among girls and activated the stereotype, regardless of condition. This explanation is supported by the consistent gender gap on the block-construction task, where girls were slower overall than boys in both elementary-school age groups. Ganley et al. (2013) have offered a similar interpretation of their results. In their experiments, they did not find stereotype threat effects on math tests in a large sample of girls in fourth through twelfth grade, but they did find consistent gender differences on two of their three experiments. Such interpretation is also consistent with previous research showing that without identity activation, women perform worse on math tests than men, and their performance can be improved by alleviating the threat, e.g. indicating that this test does not yield gender differences (Smith & White, 2002; Spencer et al., 1999).

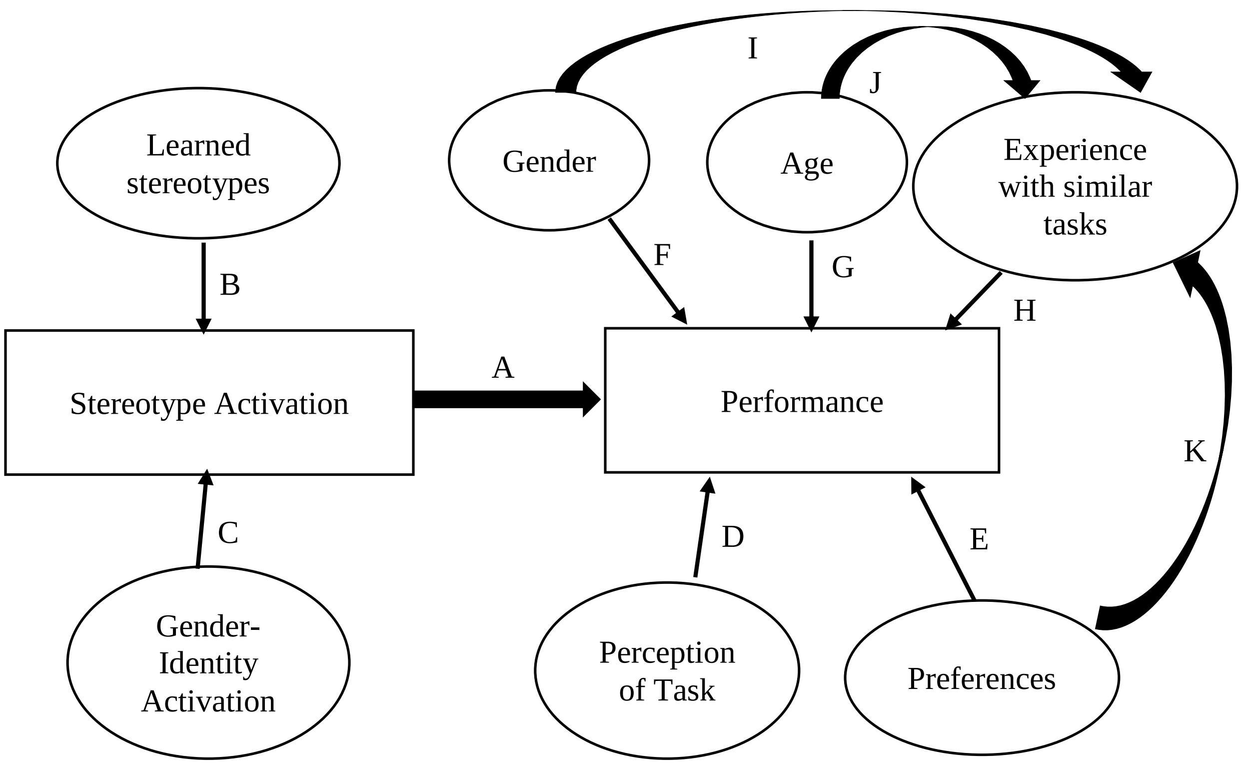

Figure 21. Stereotype Threat Process Model. Proposed model of the effect of stereotype activation on performance.

Figure 21 depicts a proposed process of stereotype threat including different variables that, based on the current findings, affect the degree to which performance is affected by the stereotype activation. The model shows that activation of a stereotype has effects on performance (A). Having learned a stereotype is a prerequisite for stereotype activation (B) and could contribute to stereotype activation without an experimental condition. In addition, activating gender identity, as in the case of Study 1, contributes to the stereotype activation (C). The extent to which performance is affected by the stereotype activation depends on a number of factors, including age, gender, experience with the task at hand, and preference for it (F, G, H, E). In addition, how participants perceive a task (D) also contributes to how stereotype activation affects performance. Experience, in turn, is affected by a participant's gender, age, and preference for this activity (I, J, K). Based on this model, girls in kindergarten through third grade performed slower overall than boys because the nature of the task activated their stereotype about LEGOs and they did not have as much experience with blocks to ward off the stress created by the activation of this negative stereotype. However, older girls' performance was comparable to boys', possibly because they had more experience with this type of task than younger girls and thus the task was not as intimidating.

Based on the model, one way to test the continuous-threat hypothesis on a block-construction task would be to alleviate girls' stress by activating a positive stereotype, for example by using pink blocks, thus altering girls' perception of the task. Results from Study 1 show that children consistently associate pink blocks with girls, when deciding who should play with blocks (Exp. 1) and who is better at building with blocks (Exp. 2). In addition, previous research using a variety of items, including toys, has found that both adults and children associate the color pink with females (Cunningham & Macrae, 2011; Ma & Woolley, 2013). If girls' performance is hampered by stereotype threat, then the use of pink blocks should cause girls to associate the task with their own gender. This could help alleviate the threat and narrow, if not eliminate, the gender gap. Such threat alleviation has been successful in previous research with adults, where participants were told that a math test does not yield gender differences (Spencer et al., 1999), thus altering the perception of the task. Likewise, in children, 11- to 13-year-old girls who were told a task measured art ability instead of geometry ability performed better when recreating a geometric shape from memory (Huguet & Regner, 2009), again altering task perception and thus blocking negative effects on performance.

However, not all stereotype threat alleviation manipulations are successful. For example, when Ganley et al. (2013) tested 13- and 14-year-old children under stereotype threat or stereotype nullification (induced by watching a video indicating either that math skill is fixed and boys are better than girls at math or that the brain is malleable and there are no gender differences) they found no effect of condition, although girls in their sample performed worse than boys overall. This inconsistency in alleviating threat, together with inconsistencies in showing stereotype threat in the first place (see Ganley et al, 2013 for review) calls into question the robustness of stereotype threat. Given the large body of research supporting the existence of stereotype threat, there is no doubt that this effect does exist. However, more thorough examination is required to systematically test what exactly evokes stereotype threat and its alleviation.

In addition to the continuous-threat explanation, there are other plausible explanations for this gender gap on the block-construction task. For example, boys may be innately better at building with blocks. Evidence of gender differences on mental rotation, a skill required for block construction, exists even in infancy (Moore & Johnson, 2008; Quinn & Liben, 2008) and supports this biological-differences hypothesis. However, research has also shown that gender differences on mental rotation among adults can be nearly eliminated by training females (Feng et al., 2007; in the short term: Miller & Halpern, 2013) or priming them with a positive stereotype (McGlone & Aronson, 2006), suggesting that learning and socialization play an important role in females' performance (Nisbett et al., 2012). Limited research with 10- and 11-year-olds has also shown that when children were trained to manually rotate images they performed better on a mental-rotation task than their performance prior to training (Wiedenbauer & Jansen-Osmann, 2008). Thus, there may be biological predispositions for superior mental rotation skills in boys, but it is also likely that experience contributes to gender differences.

Other Influences on Performance

In addition to measuring children's performance on the block-construction task, children's implicit and explicit stereotypes were assessed, and in Experiment 2, children's toy preferences were measured by administering an activities-preference measure (POAT-PM, Liben & Bigler, 2002). Parents were also asked to indicate their child's favorite toy and how frequently they played with different toys at home. The goal of these additional measures was to investigate whether there were relationships between children's attitudes, preferences, and performance. In Experiment 1, preschool girls did not express explicit stereotypical attitudes about masculine and feminine toys and activities (POAT-AM) or about who should play with blocks (except for associating pink blocks with girls). However, girls did show an implicit stereotype by being more likely to use a male pronoun to refer to a block-competition winner, similar to children's responses in Ambady et al.'s study (2001). It could be argued that children are more likely to use a male pronoun to refer to an unidentified person, regardless of a stereotype; however, studies using other implicit measures, such as the IAT, have yielded similar results in children (Galdi et al., in press). Likewise, in Experiment 2, although most children used a male pronoun to refer to the block-competition winner, this implicit stereotype did not predict speed on the block-construction task.

Unlike previous research showing that children's susceptibility to stereotype threat is related to their implicit rather than explicit stereotypes (Ambady et al., 2001; Galdi et al., in press), older children's explicit stereotypes about multicolored blocks did predict their speed on the block-construction task, such that both boys and girls were slower if they believed that boys were better at block construction. This relationship for girls is consistent with stereotype threat research. Boys being slower when they believed members of their own gender were better at this task is counterintuitive, but supported by other research. In particular, Cimpian, Mu, and Erickson (2012) found that children's performance (on an item-circling task and a mental-rotation task) was negatively affected by generic statements about boys' or girls' ability on this task, regardless of whether the statement was positive or negative. By extension, in the current study, boys who believed their group should be better at this task may have been anxious about meeting this expectation, thus performing worse than other boys.

Another factor that predicted children's speed on the block-construction task was children's preferences for and experience with LEGO blocks. How frequently children played with LEGO blocks at home and whether LEGO blocks were the child's favorite toy, as reported by parents, significantly predicted how quickly children in Exp. 2 recreated the block designs. However, this is correlational data and it is hard to tease apart innate ability and experience on this task. It could be that some children have stronger spatial abilities and skills required for block construction and thus prefer to spend time playing with LEGO blocks. Alternatively, children's skill at block construction could have been driven by their experience playing with blocks. I favor the experience interpretation given that several studies have supported the importance of early experiences on spatial-task performance (e.g., Newcombe, 1982, Sherman, 1967; see Kersh, Casey, & Young, 2008 for review). Research has also shown that training improves performance on spatial tasks in adults (Feng et al., 2007, Miller & Halpern, 2013) and children (Wiedenbauer, & Jansen-Osmann, 2008). In order to test these possibilities, children who do not spend a lot of time playing with LEGO blocks could be "trained" by spending more time playing with LEGO blocks, and then their performance could be compared to children who "naturally" like to play with LEGO blocks and to children who do not typically play with LEGO blocks. If children who have been trained perform better on block construction than in the pretest and compared to children who naturally spend more time playing with LEGO blocks, then experience is a determinant of performance. However, if their performance does not improve over pretest or compared to children who play with LEGO blocks frequently, then there may be a biological predisposition driving their performance.

Overall, the current findings present three main conclusions about stereotype threat in children: 1) girls are susceptible to stereotype threat as early as age 4, 2) block-construction speed is predicted by explicit stereotypes about blocks and preferences for or experience with blocks, and 3) girls are overall slower than boys on block construction. The first result is groundbreaking in that, to my knowledge, this is the first time that preschoolers were tested under stereotype-threat conditions and showed susceptibility to stereotype threat. By showing this susceptibility among young children, these findings highlight the danger that stereotypes present for girls in developing the cognitive skills necessary for excelling at STEM fields. In addition, by showing that children's preference for and experience with LEGO blocks predicts higher performance on the block task, the present study sheds light on the importance of encouraging children to engage in play activities that relate to STEM fields early on. Recently, there has been a surge in toys directed at girls that have typically been associated with boys, such as LEGO's new line of blocks targeting girls, like LEGO friends and LEGO Disney Princess. In fact LEGO Friends became LEGO's fourth most popular product following its launch in 2012, and despite doubling production, demand was still not met (McNally, 2013). Although toy producers’ primary interest is targeting more consumers and maximizing profit, there may be a benefit to capturing girls’ attention with toys that are known to relate to the development of mathematical skills. However, additional research, such as the proposed pink-block follow-up study, is needed to assess whether such toys can decrease stereotype threat, increase girls’ interest in such toys, and narrow the gender gap on math- and science-related tasks.

STEM Attitudes and Preferences

In addition to assessing stereotype threat, I investigated how children view women in STEM fields. Specifically, Study 2 measured children’s attitudes towards males and females employed in or learning about STEM and non-STEM fields. The results of Study 2 revealed that in general, as predicted, children in kindergarten through third grade rated women in STEM professions as less competent at their jobs than men at the same professions. Children also gave higher ratings of competence to the person of their own gender, except girls rating STEM professions, where they were more likely to rate males higher than females (especially in the older age group). Children’s tendency to favor in-groups based on gender has been supported by previous research studies (e.g., Bigler, Jones, & Lobliner, 1997); thus the in-group preference observed here is not surprising. In the case of male participants, in-group bias and STEM stereotypes both should cause boys to rate males more highly than females so it is unclear whether boys rated males highly because they viewed them as in-group members, because of their learned stereotypes, or as a result of a combination of these factors. To further explain the basis for boys’ responses, a follow-up study could ask children to rate males and females in stereotypically female professions (e.g., nurse, teacher). If boys still award higher competence ratings to males, then this means that boys have an overall tendency for in-group favoritism, regardless of field. Alternatively, if boys believe that females in such professions would be more competent than men, then they would be using their knowledge of gender stereotypes to judge individuals. There is research suggesting that children learn stereotypes about professions by the time they are in preschool (Liben et al., 2001, Liben et al., 2002, Shenouda & Danovitch, 2013), however to my knowledge no research has looked at how children evaluate a man and a woman doing the same job. Thus, the proposed follow-up study could inform this question.

Instead of attributing girls' ratings of men and women in STEM fields, which were comparable, to a competition between in-group bias and STEM stereotypes, this pattern could be attributed to girls' generally more flexible attitudes. Results from the POAT-AM measure indeed do show that boys were more stereotypical than girls, a finding mirrored by previous research (e.g., Green, Bigler, & Catherwood, 2004). To test whether girls' ratings are a result of in-group bias and stereotypes competing for their answers or whether they are just more flexible and less stereotypical, a future study could look at girls' responses to strongly feminine professions (such as nurse, teacher), in order to have stereotypes and in-group bias working in the same direction. If girls' responses strongly favor women, then girls' ratings in the current study were the result of competition between in-group bias and stereotypes. However, if they still show comparable ratings, then this would suggest that girls are overall more flexible and less stereotypical than boys.

Although children showed a discrepancy between their competence ratings of men and women in STEM professions, the points they awarded to fifth-grade boys and girls learning about STEM and non-STEM subjects, indicating how well they did on a test, appeared to be driven primarily by in-group bias (although girls ratings were similar for both males and females, suggesting that both in-group bias and stereotypes affected their ratings). There are two different interpretations for the difference in response patterns between children's judgments of adults' competence and their judgments of children's learning. First, children may be making a distinction between competence and learning and using stereotypes when judging one but not the other. If this were the case, then children could be seen as being more flexible about learning but stereotypical about adults' competence. Alternatively, *stereotype stratification* could be taking place (Steele, 2003). Stereotype stratification is when a member of a group creates a subgroup in their mind to which a stereotype does not apply. In this case children, especially girls, could be creating a subgroup labeled "girls" within the female category to whom STEM stereotypes are less likely to apply. Finding that children's stereotypical beliefs about toys and activities (POAT-AM) correlated with lower ratings for girls learning about STEM subjects but not women in STEM professions supports the stereotype stratification hypothesis. In order to better understand what is driving children's judgments of competence and learning, a follow-up study could be set-up whereby adults are described as learning about new subjects and children rate their success at learning, as in the current study. If children do not show different ratings for males and females as in the current work, then this would suggest that children distinguish between competence and learning, and do not apply stereotypes to learning. If children do show stereotypical attitudes, then this would suggest that they have been stratifying. If this were the case, then girls could potentially benefit from interventions aimed at helping them create subgroups immune to gender stereotypes, thus alleviating the need to conform to gender roles.

Children did not show gender differences in school-subject preferences, except that girls expressed liking art and music more than boys. Given the gender gap on the block-construction task in Experiment 2, children's similar school subject preferences could be pointing away from the notion that math and science attitudes develop before performance gaps, as they showed a performance gap but had similar preferences for STEM subjects. However, data about how participants were performing academically was not collected, and thus it is not possible to assess whether this sample of girls was underperforming in STEM school subjects compared to boys.

Although children's school-subject preferences did not differ by gender, there was a positive correlation between children's preferences for masculine toys and their preferences for STEM subjects. An early review of existing research at the time (Tracy, 1987) found that among children ages 3 to 13, children who had masculine toy preferences also had higher math and science grades. The current study did not collect data about participants' academic achievement, and thus it is not possible to test whether this relationship is still true. However, the association between toy preferences and STEM-subject preferences in the current study sheds light on the relationship between toys and academic achievement. There could be biological (genetic or hormonal) predispositions that make masculine toys and STEM subjects (typically associated with males) attractive to certain children, as has been suggested by several research studies (see Hines, 2011 for a review). Alternatively, it may be possible that engaging in one facilitates the other; for example, children who prefer playing with bulldozers, airplanes, and cars may develop certain spatial skills that relate to performance in STEM fields. If this were the case, then children could benefit from being encouraged to play with toys and activities that promote skills necessary to excel at STEM fields. Further research is required to identify the nature of the relationship between toy preferences and school-subject preferences.

Looking at career aspirations, Study 2 revealed that the rate at which boys and girls cited STEM career aspirations was similar, but boys were more likely to cite STEM professions typically associated with males (engineer, computer programmer), whereas girls were more likely to cite STEM professions typically associated with females (nurse, veterinarian).

Children's non-STEM career aspirations were also consistent with gender roles, where boys were more likely to cite law enforcement and athletics and girls more likely to express interest in education and arts. Thus, consistent with previous research (Etaugh & Liss, 1992), elementary school children express stereotypical career aspirations, even though they do not express differences in current school interests. This discrepancy between children's aspirations and their interests is likely to be bridged slowly as girls move away from their interest in math and science to conform to gender roles. It may be that as girls grow older they identify skills required to achieve their career aspirations. If their career aspirations do not include STEM fields, then they would be expected to lose interest in STEM subjects and possibly start feeling that they do not actually like these subjects.

Gender Similarities and Differences

In addition to the performance gap on the block-construction task discussed above, there were a number of other gender differences observed in this data. First, in relation to the block-construction task, girls rated the task as more difficult than boys did. Given that they were also slower at the task, girls' difficulty ratings may have reflected their performance. Another explanation is that girls' self-efficacy is lower than boys', such that they view tasks that have been described as difficult tasks as more difficult than boys do. Although there is consistent evidence for this explanation (Meece, 1991; Pajares & Miller, 1994; Wigfield, Eccles, & Pintrich, 1996, see Pajares & Schunk, 2001 for review), these effects typically emerge in older children.

Another gender difference observed in this study was that girls had stronger preferences for masculine toys and activities than boys did for feminine toys and activities. In addition, children were more stereotypical overall about feminine activities than masculine activities, and boys were more stereotypical than girls. These results could be explained by society's general tendency to be more accepting of girls engaged in masculine activities than boys engaged in feminine activities (e.g., Fagot, 1977). Anecdotally, as boys answered questions on POAT-PM, several of them made statements like "of course not" and "ewww" when asked how much they liked to play with baby dolls, for example. Conversely, girls did not make such exclamations when presented with masculine toys. Taken together, this suggests that girls are more flexible in their attitudes and stereotypes, possibly because they are not scrutinized as much as boys are when they engage in masculine activities.

In order not to overemphasize gender differences, it is important to also highlight gender similarities in this work. In Study 1, all but one child expressed that they liked the block-construction task. Thus, even though girls were slower than boys on the task, they expressed similar liking of the task. In Study 2, girls had similar rates of preference as boys for STEM and non-STEM school subjects and career aspirations. Also, when rating adults in non-STEM professions and rating males' and females' learning ability on STEM and non-STEM subjects, both girls and boys showed in-group favoritism. In addition, both boys and girls gave similar difficulty ratings for STEM and non-STEM professions. Looking at overall gender attitudes of boys and girls, I would argue that there are more similarities than differences (Hyde, 2005), and that it is important to address both equally, even if similarities do not generate as much interest from researchers or the public.

Developmental Trends

One of the goals of this project was to measure developmental differences in susceptibility to stereotype threat and attitudes towards STEM. Several developmental patterns emerged from this work. First, whereas preschool girls were susceptible to stereotype threat, girls in kindergarten through third grade did not show decreased performance in experimental conditions compared to the control condition. It is unlikely that stereotype threat decreases or is eliminated with age, given other research showing consistent susceptibility to stereotype threat from age 5 through adulthood (e.g., Ambady et al., 2001; Galdi et al., in press; Spencer et al., 1999). Alternatively, it may be that the block task used in the current study was easy enough that it did not threaten girl's performance by the time they were in kindergarten (as evident by the smaller gender gap in the older age group), or that the description of the task as difficult created a stressful situation for all girls in Study 2, regardless of condition (as discussed above). Because Experiment 1 included a sample of girls only, it is unclear whether preschool girls' performance would have been different from boys'. A follow-up study that includes preschool boys would be necessary to determine whether there are overall gender differences as among older children. In addition, because Experiment 1 included a relatively small sample from one Midwestern community, it would be necessary to replicate these results in other populations.

Second, in Experiment 2, overall, older children were faster at completing the block designs than younger children, as would be expected given developmental improvements in spatial and fine-motor skills. Despite this improvement in performance, there were no developmental differences in children's difficulty ratings of the block-construction task. This data is consistent with evidence that younger children are more likely to overestimate their abilities (see Bjorklund & Blasi, 2011 for review).

In Study 2, one significant developmental difference was in children's ratings of the difficulty of STEM and non-STEM professions: Older children rated STEM professions as more difficult than younger children did. In addition, older children were also more likely to rate males as more competent than females on STEM professions. This change suggests that older children have had more experience with the concept of STEM fields and understand the complexity involved in such fields. Although STEM fields may not actually be more difficult than non-STEM fields, people typically believe STEM fields are more difficult (Keil, Lockhart, & Schlegel, 2010).

Summary and Conclusion

Overall, this research elucidates how gender stereotypes can have detrimental effects on attitudes and performance in children as young as age 4. Gender stereotypes directly (via experimental condition) and indirectly (by affecting children's preferences early on) hampered children's performance on a block-construction task. Because the skills required to excel at such task are closely related to skills required to excel at math and science, this gender gap is alarming. In addition, children's stereotypical beliefs about who is more competent at STEM professions may affect their interests and preferences and how they perceive their colleagues at a later stage and contribute to the underrepresentation of women in STEM.

Broadly, this work suggests that interventions promoting STEM education should start as early as the preschool years, by exposing young children to toys and activities, such as block construction that are known to enhance skills needed for success in STEM fields. It may be beneficial to use toys that attract girls' attention (e.g., pink blocks), although more research is required to determine whether such toys can increase girls' interest in spatial activities. The data on children’s attitudes and preferences suggests that there is also a need for children to be exposed, early on, to men and women employed in STEM-related professions in order to promote these fields among children, especially girls, and to avoid having children develop gender stereotypes about STEM fields.

In conclusion, this work has highlighted important developmental trends and gender comparisons using innovative methods. By using a toy readily available to children and showing a performance gap among young children, this work has emphasized the importance of early exposure to toys that relate to success at STEM fields. By showing that children hold strong stereotypical attitudes about children and adults, the findings have underlined the importance of intervention work that attempts to eliminate such stereotypes that could have detrimental effects on children’s interests, performance, and evaluation of others. In light of these findings, several directions for future research have been identified, that could inform the debate on the origins of the gender gap in STEM fields.

APPENDICES

Appendix A

Coloring Pages

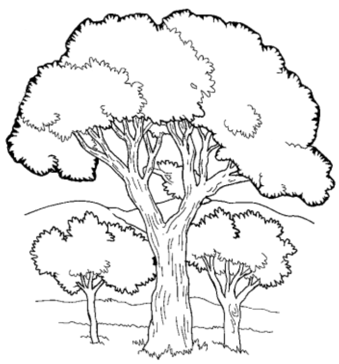

Figure 22. Coloring Page – Control Condition. Children in the control condition colored a picture of trees.

Figure 23: Coloring Pages – Feminine-Identity Activation. Girl holding a doll was used in Experiments 1 and 2, the girl only was used in Experiment 2 only.

Figure 24: Coloring Pages – Masculine-Identity Activation. These coloring pages were used in the experimental conditions in Experiment 2.

Appendix B

Experiment 1 Script

Coloring Task

Hi, my name is ______ and today I am going to give you a picture to color and then we're going to play a game with some blue blocks. I'm going to setup the blocks game now while you color the picture in this folder. This is a surprise picture for me, so just take it out of the folder, color it using these crayons, and put it back in. When I am done setting up the next game, I will come get you. Don't tell me what you colored in the picture, ok? After you finish coloring, we will play a HARD game with some blue LEGO blocks, are you ready?

Ok, it's time to play the block game now. Make sure to put the picture back in the folder, ok?

Block Task

Now I'm going to put a shape that's been built using some blocks in front of you and I am going to give you some blocks and I want you to work as fast as you can to create a shape that looks exactly like the one I show you, ok. Let's try one. (*Put Shape 0 on the table, and give child the corresponding pieces).* Can you create this shape using these blocks? *(correct child if they need help).* Ok, now you understand what you need to do.

The blocks are in two different shades of blue so make sure to match the shades, too. You are going to have 5 minutes to finish as many shapes as you can, and when you hear the beep you have to stop working on the blocks and I will count how many shapes you got correct. So you need to work as quickly as you can, but you have to get them right, too. I'm not going to tell you if you got them right, so you have to make sure you get them right. Some of the shapes will be easy and some of them will be hard. I will not give you a new shape until you tell me you are done with the one you are working on and you are ready to build the next one, ok? Again, I want you to work as quickly as you can. Are you ready?

When time is up: Time is up, now we are going to put these away so we can do some other things.

Implicit Measure

I'm going to tell you a story and I want you to pay attention because I will ask you about it later, ok?

> There were many friends that were good at this block game, but one of them stood out from the rest. This friend worked very fast with the blocks, got the shapes all right and even put together new shapes that even I couldn't do. One time, this friend entered a competition of blocks and came in first place.

Can you tell me the story again? (if they don't answer say, Remember there was a block competition, who won?)

Child referred to kid in the story as HE SHE NEUTRAL

Did you like this block game? YES NO Why/Why not?

Appendix C

LEGO Shapes

Figure 25: Shape 0. Practice shape for Experiment 1.

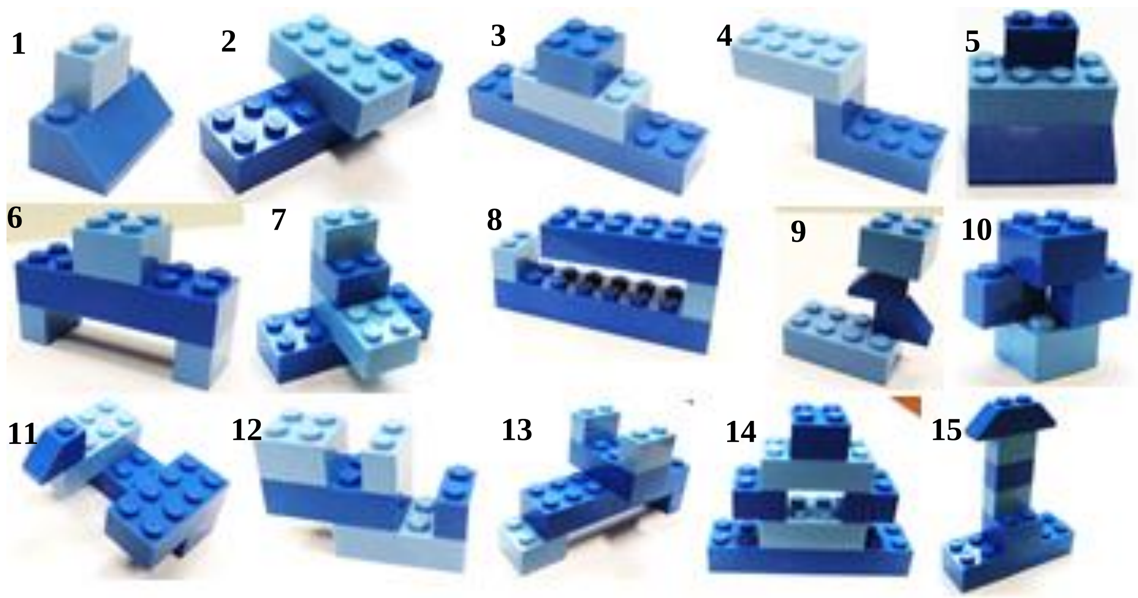

Figure 26. Block-Construction Task. In Experiment 2, shape 3 was used for practice and the eight shapes included in the task were: 4, 6, 7, 8, 10, 11, 12, and 13.

Appendix D

Experiment 2 Script

Coloring Task
Hi, my name is ______ and today I am going to give you a picture to color and then you're going to play a game with some blue blocks. I'm going to setup the blocks game now while you color the picture in this folder. So now take the picture out of the folder, color it using these crayons, and put it back in. When I am done setting up the next game, I will come get you. After you finish coloring, you will play a HARD game with some blue LEGO blocks. Are you ready?

Ok, it's time to play the block game now. Make sure to put the picture back in the folder, ok?

Block Task

Now I'm going to put a shape that's been built using some blocks in front of you and I am going to give you some blocks and I want you to work as fast as you can to make a shape that looks exactly like the one I show you. Let's try one. Can you make this shape using these block pieces? *(correct child if they need help).* Ok, now you understand what you need to do.

The blocks are in two different shades of blue so make sure to match the shades, too. You see all these different shapes on the table? When I say go, you're going to start here (point to first shape) and work as fast as you can to finish all shapes, one by one. You're going to do the same thing: look at the shape and try to make the same one using the pieces next to it. I'm going to time how fast you can do this, and after you're all done, I'm going to see how many you got right. So you have to work really fast but you also need to make sure you get them right, too. When you are done with all eight shapes, shout "done" so I can stop the time. Again, I want you to work as quickly as you can. So tell me again what you're going to do. Are you ready?

When they are done: Ok, now we are going to put these away so we can do some other things.

Implicit Measure
I'm going to tell you a story and I want you to pay attention because I will ask you about it later, ok?

There were many friends that were good at this block game, but one of them stood out from the rest. This friend worked very fast with the blocks and got the shapes all right. One time, this friend entered a block competition and came in first place.

Can you tell me the story again? *Child referred to kid as…* HE SHE NEUTRAL

Did you like this block game?	YES NO	Why/Why not?	
Was this game easy or hard?		EASY	HARD
Was it somewhat easy/hard or very easy/hard?		SOMEWHAT	VERY

Appendix E

POAT Items

Female Items
Dress up clothes
Baby dolls
Stoves & ovens
Dishes
Doll cribs
Dollhouses
Male Items
Bulldozers
Basketball
Robots
Airplanes
Lawnmowers
Dinosaurs
Neutral Items
Wooden Puzzles
Big Bouncy Balls

Table 7. POAT-AM Items. Items are presented by category. During testing, items were shuffled.

Female Items
Dress up clothes
Baby dolls
Make-up
Purses
Doll cribs
Dollhouses
Male Items
Bulldozers
Baseball
Robots
Airplanes
Cars
Dinosaurs
Neutral Items
Wooden Puzzles
Kites

Table 8. POAT-PM Items. Items are presented by category. During testing, items were shuffled.

Appendix F

Parent Questionnaire

Please complete the following questions about the child participating in this study:

Are there any other children under age 18 living full-time in this child's home? □ yes □ no

If yes, please list the ages and genders of other children living in the child's home:

1. ______ (age) M F 2. ______ (age) M F 3. ______ (age) M F
4. ______ (age) M F 5. ______ (age) M F 6. ______ (age) M F

What is your child's favorite toy? ___________________________________

Does your child have access to **LEGO blocks** at home? □ Yes □ No

If yes, how often does your child currently play with LEGO blocks?

□ All the time □ Frequently □ Sometimes □ Rarely □ Never

Does your child have access to **Dolls** at home? □ Yes □ No

If yes, how often does your child currently play with Dolls?

□ All the time □ Frequently □ Sometimes □ Rarely □ Never

Does your child have access to **Cars/Trucks** at home? □ Yes □ No

If yes, how often does your child currently play with Cars/Trucks?

□ All the time □ Frequently □ Sometimes □ Rarely □ Never

Does your child have access to **Kitchen/Tea Set** at home? □ Yes □ No

If yes, how often does your child currently play with Kitchen/Tea Set?

□ All the time □ Frequently □ Sometimes □ Rarely □ Never

Thank you!

Appendix G

Study 2 Script

Competence

Today I want to know what you think about different things.

First, I want to tell you about some grownups who work in different jobs, and I want you to tell me how good you think they are at their job. You know how different people work at different jobs, and some people are better at their job than others. We are going to use these stars to show how good they are at their jobs.

You see these 5 stars? I want you tell me how many stars you would give to each person I describe.

- 5 stars mean this person is **excellent** at the job.
- 4 stars mean this person is **good** at the job.
- 3 stars mean this person is **ok** at the job.
- 2 stars mean this person is **not so good** at the job.
- 1 star means this person is **not good at all** at the job.

(if you absolutely want to give zero stars, write the number zero next to the person, so I know you didn't just miss that).

So, for every person I tell you about, in your booklet, you're going to fill in the number of stars you think each of these grownups deserves. Ok?

So, if you think that a person is excellent at their job, how many stars would you give the person?

And if you think a person is not good at all, how many stars would you give this person?

Ok, so now I will tell you about some men and women doing different jobs and I want you to decide how many stars each of them deserves. There are no right or wrong answers. I just want to know what you think. We are not going to talk about your answers. Just color the stars in your booklet, ok? You may not know all the jobs I tell you about, but I will explain what these jobs are, and I will have some pictures to help. I want you to listen carefully to what I say before you start filling in the stars, ok?

This man and this woman are …... How good do you think each of them is at being a …………? Color the number of stars you think each of them deserves.

Biologists: A biologist knows about living things, like animals, insects, and birds. A biologist knows about where animals live, what they eat, and what sounds they make.

Advertisers: An advertiser knows all about how to get people to buy things. An advertiser makes tv commercials and signs for the streets to sell products like cars and food.

Computer Scientists: A computer scientist knows about computers. A computer scientist can fix computers and develop new programs and apps to use for work and fun.

Historians: A historian knows about what people have done in the past. A historian knows about the history of different countries, wars, and important events.

Engineers: An engineer knows about machines. An engineer can design and build new machines and can fix machines and make them work better.

Sociologists: A sociologist knows about how different people in different countries live. A sociologist studies how different people do things and what they believe.

Mathematicians: a mathematician knows about numbers and counting. A mathematician can count very large numbers very quickly and can count a lot of things like money and people.

Journalists: A journalist knows about recent news. A journalist reports about important news so that people know what is happening in their area and in other states and countries.

Ok, we're done with the first part. Now let's stretch our arms really wide. Great job.

Future Aspiration

Now let me ask you a quick question. Don't say your answer out loud, just flip the page and write your answer. Raise your hand if you need help spelling that. What would you like to be when you grow up? It doesn't have to be anything we discussed today.

Job Difficulty

You know how some things are easy to do and some things are hard to do? I'm going to tell you about jobs that we just talked about, and I want you to tell me how hard you think these jobs are for grownups. We will use the numbers 1 to 5 to help.

1 – means this job is very easy

2 – means this job is kind of easy

3 – means this job is medium (neither easy nor hard)

4 – means this job is kind of hard

5 – means this job is very hard

So if you think that a job is very hard for grownups, which number would you circle? (make sure they say 5). And if you think this job is very easy for grownups, which number would you circle? (make sure they say 1).

Ok, let's start. Again, no talking, just circle in your booklet. How hard do you think it is to be a….. for a grownup?

- Biologist
- Advertiser
- Computer Scientist
- Historian
- Engineer
- Sociologist
- Mathematician
- Journalist

Learning Ability

Ok, now we are going to do something a little different. You know how some students like things they learn at school more than others, and some learn faster than others. I am going to tell you about some fifth-grade boys and girls who learned all about a new, hard topic, and then took a test about that topic. The test had 5 questions. I want you to decide how many questions you think each of them got right on the test they took. We will use the numbers 1-5 again, but this time they will mean different things.

5 – means the student got all 5 questions right

4 – means the student got 4 questions right

3 – means the student got 3 questions right

2 – means the student got 2 questions right

1 – means the student got 1 questions right

So, you will circle the number in your booklet.

So, if you think a kid got all questions right, which number would you circle? (make sure they say 5)

And if you think this kid got only 1 question right, which number would you circle? Remember, there are no right or wrong answers, I just want to know what you think. Again, we are not going to talk, just circle in your booklet, ok? Am going to tell you what they learned about and explain what this really is, in case you don't know. Please listen to what I say before you start circling.

This boy and this girl took a hard class to learn all about advanced (name of subject). They learned all about (use definition below). Then they took a hard test. The test had 5 questions, how many do you think each of them got right? Circle how many you think each of them got right.

Math: learned all about algebra and geometry including fractions and learned how to calculate the area and surface of different shapes.
Music: Learned all about different composers and different styles of music and learned how to play music on the piano.
Life Science: learned all about different organs of the human body and different parts of the brain and what their functions are.

Art: learned all about different styles of painting and about different artists and learned how to accurately draw a nature scene.
Physical Science: learned all about what materials are made of and learned how to combine different materials to create new things.
Social Studies: learned all about the history of different countries and learned about the locations of many places in the world.
Computers: learned all about how computers work and the different parts inside a computer and learned how to fix a computer when it's not working.
Grammar: learned all about how to form long sentences, how to spell different words, and how to use punctuation correctly.

School subject preferences

Now we're going to do something a little different. I'm going to ask you how much do you like things that you learn at school. You see these faces? I want you to use the face that best describes how much you like each thing I mention.

 Means you like it a lot.

 Means you like it.

 Means It's ok.

 Means you don't like it a lot.

 Means you don't like it at all.

So I want you to circle the one that best describes how much you like something. So let's say you like something a lot, which one would you circle? Let's say you don't like something at all, which one would you circle? Again, there are no right or wrong answers, I just want to know what you think. And no talking, just circle in your booklet.

How much do you like:

- Math
- Spelling
- Art
- Science
- Reading
- Music
- Computers

Which one is your favorite from this list? You can circle your favorite. Circle only 1. Call me if you need help reading the words.

Ok, we're all done. You all did a great job. Thank you!

Appendix H

Study 2 Sample Pages

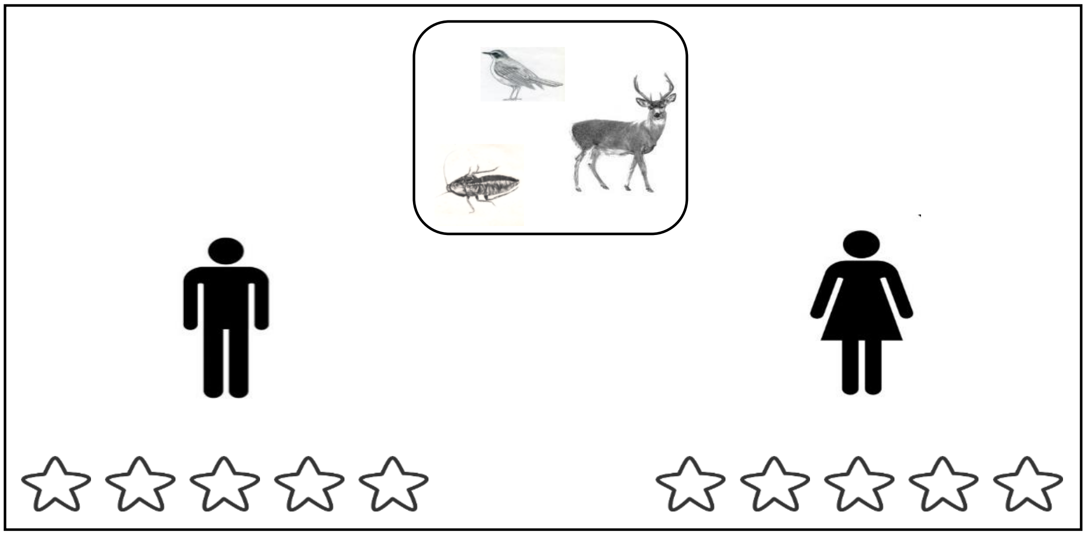

Figure 27. Sample Competence Ratings Page. Each profession was presented on a separate sheet. This sample represents biologist.

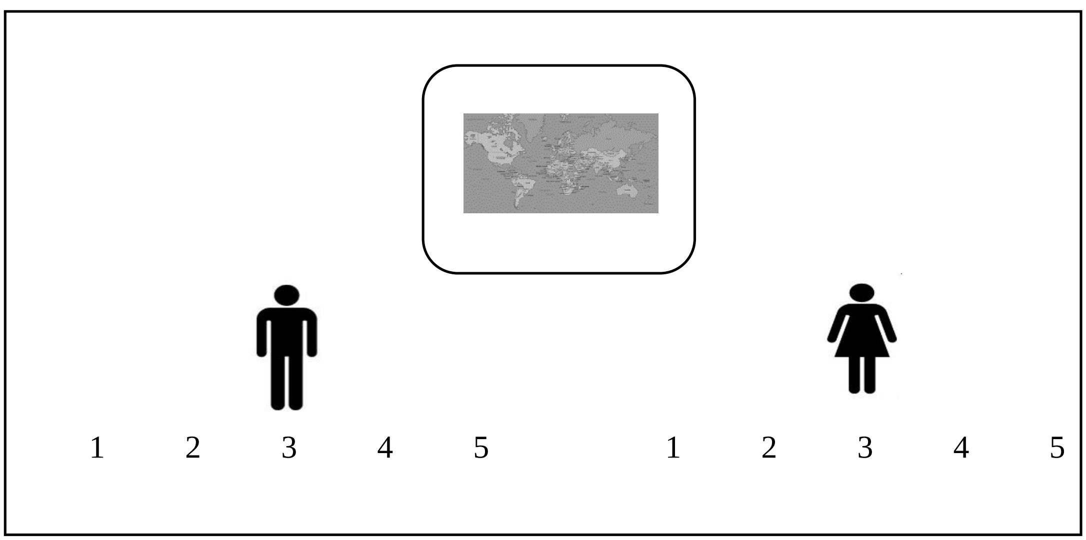

Figure 28. Sample Learning-Ability Ratings Page. Each subject was presented on a separate sheet. This sample represents social studies.

Appendix I

Study 2 Individual Items

	Overall (N = 134)		Boys (n = 70)		Girls (n = 64)	
	M	SD	M	SD	M	SD
STEM Professions						
Male Mathematician	4.14	1.29	4.36	1.14	3.91	1.41
Female Mathematician	3.86	1.48	3.60	1.63	4.14	1.26
Male Engineer	4.09	1.43	4.44	1.09	3.70	1.65
Female Engineer	2.83	1.77	2.74	1.84	2.92	1.70
Male Computer Scientist	4.25	1.24	4.41	1.26	4.08	1.21
Female Computer Scientist	3.77	1.48	3.39	1.64	4.19	1.15
Male Biologist Rating	4.01	1.41	4.19	1.41	3.81	1.39
Female Biologist	3.69	1.56	3.51	1.74	3.89	1.32
Non-STEM Professions						
Male Journalist	3.71	1.41	3.87	1.33	3.53	1.49
Female Journalist	3.60	1.49	3.33	1.60	3.89	1.31
Male Sociologist	3.57	1.45	3.60	1.43	3.55	1.49
Female Sociologist	3.48	1.68	3.04	1.78	3.95	1.43
Male Historian	3.60	1.53	3.59	1.53	3.63	1.54
Female Historian	3.47	1.62	3.11	1.74	3.86	1.38
Male Advertiser	3.61	1.58	3.94	1.40	3.25	1.69
Female Advertiser	3.41	1.65	2.88	1.71	3.98	1.39

Table 9. Competence Ratings: Individual Items. Children's competence ratings for males and females in STEM and non-STEM professions

	Overall		Boys		Girls	
	M	SD	M	SD	M	SD
STEM Subjects						
Boy learning Computers	3.57	1.53	3.63	1.52	3.52	1.55
Girl learning Computers	3.11	1.55	2.71	1.62	3.53	1.37
Boy learning Physical Sciences	3.63	1.39	3.86	1.31	3.38	1.45
Girl learning Physical Sciences	3.04	1.51	2.75	1.49	3.34	1.49
Boy learning Life Sciences	3.35	1.53	3.53	1.61	3.14	1.41
Girl learning Life Sciences	3.09	1.60	2.77	1.70	3.44	1.41
Boy learning Math	3.55	1.53	3.93	1.38	3.14	1.59
Girl learning Math	3.38	1.59	2.89	1.66	3.94	1.32
Non-STEM Subjects						
Boy learning Grammar	3.76	1.48	3.87	1.64	3.63	1.27
Girl learning Grammar	3.54	1.54	3.14	1.70	3.97	1.22
Boy learning Social Studies	3.75	1.33	4.04	1.18	3.44	1.42
Girl learning Social Studies	3.35	1.47	3.13	1.47	3.59	1.43
Boy learning Art	3.59	1.47	3.77	1.35	3.39	1.58
Girl learning Art	3.98	1.47	3.57	1.68	4.42	1.05
Boy learning Music	3.44	1.46	3.67	1.48	3.19	1.41
Girl learning Music	3.69	1.53	3.49	1.65	3.92	1.37

Table 10. Learning Ratings: Individual Items. Children's ratings of boys' and girls' learning about new subjects. Values are out of 5, where 1 = boy/girl answered one question correctly and 5 = boy/girl answered five questions correctly.

Appendix J

Career Aspirations

STEM	Non-STEM	
scientist (6)	teacher (14)	animal rescuer
vet (6)	police officer (8)	zoo keeper (4)
doctor (5)	soldier/army man (5)	farmer (4)
paleontologist (3)	spy (2)	sell animals
dentist (2)	FBI agent (2)	star/model (2)
nurse	firefighter (4)	artist/painter (3)
astronaut (2)	professional/olympics athlete (3)	fashion designer
pilot (2)	football player (2)	hair cutter
computer designer/programmer (2)	dirt bike driver	LEGO designer
video game maker	baseball player (2)	librarian
engineer	basketball player (2)	musician
geologist	cheerleader/cheerleading coach (2)	news man
marine biologist	dance teacher (2)	nun
mechanic	dancer (2)	work at daycare/babysitter (3)
	golfer	owner of toy store
	gymnastics star	painter
	jockey	president
	karate champion	principal
	wwe referee/wrestler (2)	server
	soccer player	taste tester
	adventurer/explorer/hunter (5)	wiggle master
	superhero	work at church
	american ninja warrior (2)	millionaire
		mom

Table 11. Children's Career Aspirations. Children's career aspirations by STEM and non-STEM. Frequency of each item in brackets.

REFERENCES

REFERENCES

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science*, 12, 385–390.

Aronson, J., & Good, C. (2003). The development and consequences of stereotype vulnerability in adolescents. In F. Pajares & T. Urdan (Eds.), *Adolescence and education*: *Vol. 2. Academic motivation of adolescents* (pp. 299–330). Greenwich, CT: Information Age.

Aronson, J., & McGlone, M. (2009). Stereotype and social identity threat. In T. Nelson (Ed.), *The handbook of prejudice, stereotyping, and discrimination*. New York: Guilford.

Ball, L. C., Cribbie, R. A., Steele, J. R. (2013). Beyond gender differences: Using tests of equivalence to evaluate gender similarities. *Psychology of Women Quarterly*, 37 (2), 147-154.

Baron-Cohen, S. (2004). *Essential Difference: Male and Female Brains and the Truth about Autism*. New York: Basic Books.

Barres, B. A. (2006). Does gender matter?. *Nature*, *442*(7099), 133-136.

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

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, *136*, 256–276.

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

Bigler, R. S., Jones, L. C., & Lobliner, D. B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child Development*, *68*(3), 530-543. doi: 10.1111/j.1467-8624.1997.tb01956.x

Bjorklund, D., & Blasi, C. H. (2011). *Child and Adolescent Development: An Integrated Approach*. Cengage Learning.

Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science, 12*(3), 225-229.

Bodenhausen, G. V. (1990). Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychological Science, 1*(5), 319-322.

Brehm, J.W. (1999). The intensity of emotion. *Personality and Social Psychology Review, 3,* 2-22.

Brosnan, M. J. (1998). Spatial ability in children's play with Lego blocks. *Perceptual and Motor Skills, 87*, 19-28.

Caldera, Y. M., Mc Culp, A., O'Brien, M., Truglio, R. T., Alvarez, M., & Huston, A. C. (1999). Children's play preferences, construction play with blocks, and visual-spatial skills: Are they related?. *International Journal of Behavioral Development*, *23*(4), 855-872.

Cheng, Y.L. and Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development*, 2-11.

Cimpian, A., Arce, H.-M. C., Markman, E. M., & Dweck, C. S. (2007). Subtle linguistic cues affect children's motivation. *Psychological Science,* 18, 314–316. doi:10.1111/j.1467-9280.2007.01896.x.

Cimpian, A., Mu, Y., & Erickson, L. C. (2012). Who is good at this game? Linking an activity to a social category undermines children's achievement. *Psychological Science*, *23*(5), 533–541.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155.

Cohen J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37-46.

Conry-Murray, C., and Turiel, E. (2012). Jimmy's baby doll and Jenny's truck: Young children's reasoning about gender norms. *Child Development, 83* (1), 146-158. doi:10.1111/j.1467-8624.2011.01696.x

Cundiff, J. L., Vescio, T. K., Loken, E., & Lo, L. (2013). Do gender–science stereotypes predict science identification and science career aspirations among undergraduate science majors?. *Social Psychology of Education*, *16*(4), 541-554.

Cunningham, S. J. & Macrae, C. N. (2011). The colour of gender stereotyping. *British Journal of Psychology*, *102*, 598-614.

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-gender stereotypes in elementary school children. *Child Development*, *82*, 766–779.

Cvencek, D., Meltzoff, A. N., & Kapur, M. (2014). Cognitive consistency and math–gender stereotypes in Singaporean children. *Journal of Experimental Child Psychology*, *117*, 73-91.

Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology*, 88, 276–287.

del Río, M. F., & Strasser, K. (2013). Preschool children's beliefs about gender differences in academic skills. *Sex Roles*, *68*(3-4), 231-238.

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly,* 11, 135–172.

Eccles, J. S. (1989). Bringing young women to math and science. In *Gender and Thought: Psychological Perspectives* (pp. 36-58). New York: Springer.

Else-Quest, N. M., Mineo, C. C., & Higgins, A. (2013). Math and science attitudes and achievement at the intersection of gender and ethnicity. *Psychology of Women Quarterly*, *37*(3), 293-309.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999).Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology*: General, 128, 309–331.

Entwistle, D. R., Alexander, K. L., Pallas, A. M., & Cardigan, D. (1987). The emergent academic self-image of first graders: Its response to social structure. *Child Development*, 58, 1190–1206.

Etaugh, C., & Liss, M. B. (1992). Home, school, and playroom: Training grounds for adult gender roles. *Sex Roles*, 26(3-4), 129-147.

Fagot, B. I. (1977). Consequences of moderate cross-gender behavior in preschool children. *Child Development*, 902-907.

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, *18*(10), 850-855.

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology*, *38*(4), 519-533.

Freeman, G.H. and Halton, J.H. (1951). Note on exact treatment of contingency, goodness of fit and other problems of significance. *Biometrika,* 38:141-149.

Frome, P.M., Alfeld, C.J., Eccles, J.S., & Barber, B.L. (2006). Why don't they want a male-dominated job? An investigation of young women who changed their occupational aspirations. *Educational Research and Evaluation, 12*(4), 359-372.

Galdi, S., Cadinu, M. & Tomasetto, C. (in press). The roots of stereotype threat: When automatic associations disrupt girls' math performance. *Child Development*, manuscript accepted for publication (01/03/2013).doi:10.1111/cdev.12128

Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype threat effects on girls' mathematics performance. *Developmental psychology*, *49*(10), 1886.

Ganley, C. M., Vasilyeva, M., & Dulaney, A. (in press). Spatial Ability Mediates the Gender Difference in Middle School Students' Science Performance. *Child development*.

Gilbert. D. T. & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology 60*, 509-517.

Goble, P., Martin, C. L., Hanish, L. D., & Fabes, R. A. (2012). Children's gender-typed activity choices across preschool social contexts. *Sex Roles, 67*, 435–451

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464-1480.

Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2011). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles*, this issue. doi:10.1007/s11199-011-9996-2.

Hahn, N., Heil, M., & Jansen, P. (2010). Preschoolers' mental rotation: Sex differences in hemispheric asymmetry. *Journal of Cognitive Neuroscience*, 22, 1244−1250.

Halpern, D. F., Benbow, C., Geary, D., Gur, D., Hyde, J., & Gernsbacher, M. A. (2007). The science of sex-differences in science and mathematics. *Psychological Science in the Public Interest*, 8, 1–52. doi:10.1111/j.1529-1006.2007.00032.x

Hamilton, D.L., & Trolier, T.K. (1986). Stereotypes and stereotyping: An overview of the cognitive approach. In J. Dovidio & S. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 127-163). Orlando, FL: Academic Press.

Handelsman J, et al. (2005) Careers in science. More women in science. *Science* 309:1190–1191.

Herbert, J. & Stipek, D. T. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Journal of Applied Developmental Psychology*, *26*(3), 276-295.

Hill, Catherine, Christianne Corbett, Andresse St. Rose. (2010). *Why So Few? Women In Science, Technology, Engineering, And Mathematics.* Washington, D.C.: AAUW.

Hines, M. (2011). Gender development and the human brain. *Annual Review of Neuroscience, 34*, 69-88.

Howell, D. C. (2002) Statistical Methods for Psychology, 5th ed.. Belmont, CA: Duxbury Press.

Huguet, P., & Regner, I. (2009). Counter-stereotypic beliefs in math do not protect school girls from stereotype threat. *Journal of Experimental Social Psychology, 45*(4).

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics, 1*(1), 69-82.

Hyde, J. S. (2005). The gender similarities hypothesis. *American psychologist*,*60*(6), 581-592.

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect. *Psychology of women Quarterly, 14*(3), 299-324.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C.C. (2008). Gender similarities characterize math performance. *Science*, 321, 494–495. doi:10.1126/science.1160364.

Inzlicht, M., McKay, L., & Aronson, J. (2006). Stigma as ego depletion: How being the target of prejudice affects self-control. *Psychological Science, 17*, 262–269.

Jacobs, J. E. (1991). The influence of gender stereotypes on parent and child math attitudes: Differences across grade-levels. *Journal of Educational Psychology, 83*, 518-527.

Jacobs, J. E., & Bleeker, M. M. (2004). Girls' and boys' developing interests in math and science: Do parents matter?. *New Directions for Child and Adolescent Development*, *2004*(106), 5-21.

Jahoda, G. (1979). ON the nature of difficulties in spatial-perceptual tasks: Ethnic and gender differences. British Journal of Psychology, 70, 351-363.

Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology, 93*, 544–564.

Kawakami, K., Steele, J. R., Cifa, C., Phills, C. E., & Dovidio, J. F. (2008). Approaching math increases Math = Me, Math = Pleasant, and perseverance at math in women. *Journal of Experimental Social Psychology*, 44(3), 818-825.

Keil, F. C., Lockhart, K. L., & Schlegel, E. (2010). A bump on a bump? Emerging intuitions concerning the relative difficulty of the sciences. *Journal of Experimental Psychology: General*, *139*(1), 1-15.

Kersh, J., Casey, B. M., & Young, J. M. (2008). Research on spatial skills and block building in girls and boys. *Contemporary perspectives on mathematics in early childhood education*, 233-251.

Liben, L. S., & Bigler, R. S. (2002). The developmental course of gender differentiation: Conceptualizing, measuring, and evaluating constructs and pathways. *Monographs of the Society for Research in Child Development, 67, (2, Serial No. 269).*

Lubinski, D. (2010). Spatial ability and STEM: a sleeping giant for talent identification and development. *Personality and Individual Differences*, 49, 344-351.

Lytton, H., & Romney, D. M. (1991). Parents' differential socialization of boys and girls: A meta-analysis. *Psychological Bulletin*, *109*(2), 267.

Ma, L., & Woolley, J. (2013). Young Children's Sensitivity to Speaker Gender When Learning from Others. *Journal of Cognition and Development,* 14(1), 100-119. doi:10.1080/15248372.2011.638687

Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology, 66*, 37-7.

McGlone, M. S. & Aronson, J. (2007). Forewarning and forearming stereotype-threatened students. *Communication Education, 56*, 119-133.

McNally, M. (2013). Successful LEGO strategy delivers continued strong growth. Retrieved April 8, 2014 from http://aboutus.lego.com.

Meece, J. L. (1991). The classroom context and students' motivational goals. In M. L.Maehr & P. R. Pintrich (Eds.), *Advances in Motivation and Achievement* (Vol. 7, pp. 261-285). Greenwich, CT: JAI Press.

Miller, D. I., & Halpern, D. F. (2013). Can spatial training improve long-term outcomes for gifted STEM undergraduates?. *Learning and Individual Differences, 26*, 141-152.

Mix, K. S. and Cheng, YL. (2012). The relation between space and math: Developmental and educational implications. *Advances in Child Development and Behavior*, 42, 197-243.

Moè, A. (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. *Learning and Individual Differences*, 19, 21−27.

Moè, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual Differences*, 16, 369−377.

Moore, D. S., & Johnson, S. P. (2008). Mental Rotation in Human Infants A Sex Difference. *Psychological Science, 19*(11), 1063-1066.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, *109*(41), 16474-16479.

National Science Foundation, National Center for Science and Engineering Statistics. 2013. *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2013*. Special Report NSF 13-304. Arlington, VA. Available at http://www.nsf.gov/statistics/wmpd/.

Neuville, E., & Croizet, J.-C. (2007). Can salience of gender identity impair math performance among 7-8 years old girls? The moderating role of task difficulty. *European Journal of Psychology of Education, 22*(3), 307-316.

Newcombe, N. (1982). Gender-related differences in spatial ability: Problems and gaps in current approaches. In M. Poegal (Ed.), *Spatial Abilities.* New York: Academic Press.

Newcombe, N. S. (2010). Picture this: increasing Math and Science Learning by improving Spatial Thinking. *American Educator*, summer 2010, 29-43.

Nisbett, R. E. (2009). Intelligence and how to get it: Why schools and cultures count. New York: Norton.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: new findings and theoretical developments. *American Psychologist*, 67(2), 130.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002b). Math= male, me= female, therefore math≠ me. *Journal of Personality and Social Psychology*, *83*(1), 44-59.

Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002a). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*(1), 101.

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin*,*29*(6), 782-789.

Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement?. *Contemporary Educational Psychology*,*26*(3), 291-310.

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

Pajares, F., & Schunk, D. (2001). The development of academic self-efficacy. In Wigfield & J. Eccles (Eds.), *Development of achievement motivation.*

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

Pronin, E., Steele, C. M., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: Women and mathematics. *Journal of Experimental Social Psychology, 40,* 152–168.

Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science,* 19, 1067–1070. doi:10.1111/j.1467-9280.2008.02201.x

Reilly, D., & Neumann, D. L. (2013). Gender-Role Differences in Spatial Ability: A Meta-Analytic Review. *Sex roles*, *68*(9-10), 521-535.

Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language*, *39*, 418–436.

Ruble, DN.; Martin, CL.; Berenbaum, SA. (2006). Gender development. In: Eisenberg, N., (Ed.) *Handbook of Child Development*. New York: Wiley.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*, 440–452.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*, 336–356.

Seibt, B., & Förster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology,* 87, 38–56.

Serbin, L. A., Poulin-Dubois, D., & Eichstedt, J. A. (2002). Infants' response to gender-inconsistent events. *Infancy, 3*, 531 – 542.

Shenouda, C. & Danovitch, J. H. (2013). Does a male nurse know about football?: American and Egyptian preschoolers' understanding of gender and expertise. *Journal of Cognition and Culture, 13*(3-4): 231-254.

Shenouda, C. K. & Danovitch, J. H. (in prep.) Who knows best? Preschooler's trust in the testimony of male and female informants.

Sherman, J. A. (1967). Problem of gender differences in space perception and aspects of intellectual functioning. *Psychological Review, 74,* 290-299.

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

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development*, 973-980.

Signorella, M. L., Bigler, R. S., & Liben, L. S. (1993). Developmental differences in children's gender schemata about others: A meta-analytic review. *Developmental Review, 13*(2), 147-183.

Smith, J. L., & White, P. H. (2002). An examination of implicitly activated, explicitly activated, and nullified stereotypes on mathematical performance: It's not just a woman's issue. *Sex Roles*, 47, 179–191. doi:10.1023/A:1021051223441

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science?: a critical review. *American Psychologist*, *60*(9), 950-958.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, 4–28. doi:10.1006/jesp.1998.1373.

Steele, C.M., & Aronson, J. (1995). Stereotype Threat and the intellectual test-performance of African-Americans. *Journal of personality and Social Psychology*, *69* (5): 797-811

Steele, J. R., & Ambady, N.(2006). "Math is hard!": The effect of gender priming on women's attitudes. *Journal of Experimental Social Psychology*, 42(4), 428-436.

Steele, J.(2003). Children's gender stereotypes in mathematics: The role of stereotype stratification. *Journal of Applied Social Psychology*, 33(12), 2587-2606.

Steele, J.R., Reisz, L., Williams, A., & Kawakami, K. (2007). Women in mathematics: Examining the hidden barriers that gender stereotypes can impose. In R. Burke & M. Mattis (Eds.), *Women and minorities in science, technology, engineering and mathematics: Upping the numbers* (pp.159-183). London: Edward Elgar.

Summers, L. (2005). Remarks at NBER conference on diversifying the science and engineering workforce. Retrieved March 25, 2006 from http://www.president.harvard.edu/speeches/2005/nber.html.

Tracy, D. M. (1987). Toys, spatial ability, and science and mathematics achievement: Are they related?. *Sex Roles*, *17*(3-4), 115-138.

Tukey, J. W. (1977). Exploratory data analysis. *Massachusetts: Addison-Wesley*.

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., Newcombe, N. S., Filipowicz, A. T., & Chang, A. (2014). Deconstructing Building Blocks: Preschoolers' Spatial Assembly Performance Relates to Early Mathematical Skills. *Child Development, 85* (3), 1062–1076.

Wiedenbauer, G., & Jansen-Osmann, P. (2008). Manual training of mental rotation in children. *Learning and instruction*, 18(1), 30-41.

Wigfield, A., Eccles, J. S., & Pintrich, P. R. (1996). Development between the ages of 11 and 25. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 148-185). New York: Macmillan.

Wolfgang, C. H., Stannard, L. L., & Jones, I. (2001). Block play performance among preschoolers as a predictor of later school achievement in mathematics. *Journal of Research in Childhood Education*, 15, 173-180.

Xie, Y., & Shauman, K. (2003). *Women in science: Career processes and outcomes*. Cambridge, MA: Harvard University.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*,*18*(5), 459-482.

Zhang, X., Koponen, T., Räsänen, P., Aunola, K., Lerkkanen, M. K., & Nurmi, J. E. (2014). Linguistic and spatial skills predict early arithmetic development via counting sequence knowledge. *Child Development*, *85* (3), 1091–1107.