THE UNIVERSITY OF CHICAGO

LEARNING UNDER PRESSURE: STEREOTYPE THREAT AND EVALUATIVE PERFORMANCE PRESSURE IN THE MATHEMATICS CLASSROOM

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPARATIVE HUMAN DEVELOPMENT

BY

EMILY LYONS

CHICAGO, ILLINOIS

JUNE 2019

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

There are a great many people whose support, ideas and inspiration were integral to this work and whom I would like to thank.

To start, I would like to thank the students, teachers and principals who participated in this research. This work, in a very real way, would not have been possible without your willingness to open your classrooms and to participate in this study. In addition to the students who participated in this research itself, I would like to thank my former high school students. The insights and experiences you shared with me, particularly as we worked to prepare for the Physics Regents, shaped this research and are insights I often return to when attempting to make sense of experimental findings.

I would like to thank my committee members. Lindsey Richland, Margaret Beale Spencer and Susan Levine, thank you for offering your invaluable time, feedback and support. I am grateful for the different expertise each of you brought to this project from its earliest stages up to the final hearing. These different perspectives pushed me to be a better researcher and to take ownership of this work. I feel incredibly lucky and privileged to have had the opportunity to work with and learn from each of you. Margaret Beale Spencer, thank you for challenging me to engage deeply with theoretical frameworks and to locate my work within a particular context. Guanglei Hong, thank you for reading my work closely, pushing me to always consider alternative explanations for findings, and challenging me to understand the benefits and tradeoffs of different statistical methodologies. Susan Levine, thank you for the insights, new ideas and enthusiasm you offered throughout this work.

Thank you especially to my advisor, Lindsey Richland. I am grateful for the close feedback you have provided on my writing, for how you have pushed and supported me to move outside my comfort zone and take on new challenges as a grad student such as organizing symposia, and for your genuine interest and enthusiasm for discussing new research ideas. I have learned so much from you and feel incredibly lucky to have you as an advisor.

I would also like to thank the members of my cohort and the many amazing and talented members of the Learning Lab, who have contributed to this project and who have been my friends and colleagues throughout my graduate studies. I would like to especially thank Nina Simms, who in the first years of my graduate studies took the time to teach me many invaluable research skills- from designing posters to tips and tricks of data cleaning and analysis. Thank you especially as well to Rebecca Frausel for your advice, feedback, friendship and encouragement throughout my graduate studies. I am very grateful to you for sharing tips and advice as you navigated many hurdles of grad school one year ahead of me. Thank you especially as well to Ellen Klostermann Wallace for your ideas, feedback and support. Finally, thank you to the many hardworking and talented Research Assistants who made it possible to complete these studies, especially Tyler Warner, Dana Stradley and Apporva Shivaram. Without you this work would not have been possible.

I would also like to acknowledge sincere gratitude for support from the National Science Foundation and from the National Academy of Education/Spencer foundation, which funded this dissertation.

Finally, thank you most especially to my family.

To my parents, thank you your constant support, love, and encouragement throughout my graduate studies. Thank you for raising my sisters and I to develop a sense of inquisitiveness and possibility, to desire to do work with meaning and to always push ourselves to do our best work. Thank you most especially for setting an example of hard work, commitment, fun, silliness, and caring in our home that I strive to emulate. Thank you to my sisters, Claire, Abby and Genny, for always showing interest in my work, for your encouragement throughout my graduate studies, and for being awesome, amazing and inspirational people who in addition to being my sisters are my best friends.

Thank you to my husband, Andy, for your love, support, caring and patience. Thank you for always being willing to listen to me vent and for your enthusiasm about talking through questions about statistical analyses. Most importantly, thank you for always being there for me. I could not have completed this work without you. Owen, you too contributed to the successful completion of this work. Having you to return to each day helped me to avoid procrastination and taught me to value my time. Most importantly, Owen and Andy, thank you for the joy and meaning you have brought to my life.

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

My dissertation examines how experiencing heightened pressure during conceptual mathematics instruction impacts children's mathematics learning and engagement. Classrooms can be stressful places for many students, with the pressures of children's larger socio-cultural contexts often taking shape in everyday interactions. In the current educational landscape, many students feel a great deal of pressure at school and may also worry that their academic abilities will be judged based on negative stereotypes, a phenomenon often referred to as *stereotype threat* (see Steele & Aronson, 1995). These pressures can be especially pronounced in the mathematics classroom: a "high stakes" subject area in which stereotypes remain highly salient. In a series of classroom-based experiments, I compare impacts of two different but sometimes co-occurring sources of pressure that many students experience in the mathematics classroom: stereotype threat (pressure from increased salience of a self-relevant negative stereotype, and the potential of being judged stereotypically) and evaluative performance pressure (pressure from the possibility of obtaining or losing an incentive). While both sources of pressure can increase anxiety, they differ in the extent to which identity is implicated and threatened. Although most prior research on pressure and academic achievement has focused exclusively on testing situations, findings from these experiments indicate that the role of pressure in shaping academic achievement extends beyond impacts on test performance to also shape initial knowledge acquisition. Impacts of pressure on student learning depended on both the pressure source and student characteristics.

CHAPTER ONE: INTRODUCTION

Classrooms can be busy, bustling, chaotic, stressful places, with the social pressures of children's larger socio-cultural context sometimes taking shape in everyday interactions. In the current competitive educational landscape, many children feel a great deal of pressure at school (e.g. Luthar & Kumar, 2018; Watson, Johanson, & Dankiw, 2014). Coupled with this general evaluative pressure, children from marginalized groups may also worry that their academic abilities will be judged based on negative stereotypes (e.g. Larnell, Boston & Bragelman, 2014; Legette, 2018; Nasir et al., 2009, 2017). Thus, children may experience both generalized evaluative performance pressure as well as stereotype threat in the mathematics classroom, worrying that their performance will be judged through the lens of negative stereotypes (Steele & Aronson, 1995). These pressures can be especially pronounced in mathematics contexts, a high stakes subject area in which stereotypes are often particularly salient (Leslie, Cimpian, Meyer, & Freeland, 2015; Cimpian, & Leslie, 2017).

How do these pressures shape children's learning and engagement in mathematics? While pressure can in some cases increase motivation and effort (e.g. Angrist & Lavy, 2009; Levy, List & Sadoff, 2016), it can also result in distracting thoughts and worries that interfere with cognition (see Maloney, Sattizahn, & Beilock, 2014). Evaluative performance pressure and stereotype threat share some commonalities. Importantly, both sources of pressure can lead to intrusive thoughts and worries that tax executive functions, (EFs), which are cognitive resources needed for attentional control, manipulation of mental representations, and task switching (Miyake et al., 2000; Schmader & Johns, 2003; Schmaeder & Beilock, 2012). However, the extent to which identity is explicitly implicated and potentially threatened differs, which could mean that these sources of pressure have different implications for children's mathematics trajectories.

In a series of classroom-based experiments, my dissertation examines how heightened pressure from either or both stereotype threat and evaluative performance pressure impacts early adolescents' learning and engagement when experienced during initial mathematics learning opportunities. While most prior research on stereotype threat, pressure and academic achievement has focused exclusively on testing situations, many students also feel a great deal of pressure during everyday math instruction. This could have far-reaching consequences, impacting not only performance but initial knowledge formation as well. As learning is built on a foundation of prior knowledge, failure to master a concept at one time point could also impact subsequent learning opportunities, and decreased enjoyment and identification within a domain could have long-range impacts on course selection, academic identities, career goals and persistence.

I focus on early adolescents because, due to both inherent developmental processes and the particular structure of schooling that predominates in the US, experiences and outcomes during early adolescence are pivotal in shaping academic trajectories (see Swanson, Spencer & Petersen, 1998; Spencer & Swanson, 2013). Cognitive changes occurring during adolescence increase young people's capacity for complex thought and social comparisons- but also heighten vulnerability to negative feedback from the environment. As identity construction and developing a sense of efficacy are keys tasks of adolescence, experiences during this developmental period can instigate recursive processes that have long-range implications for academic identities and aspirations (Erikson, 1959; Marcia, 1980; Swanson, Spencer & Petersen, 1998; Spencer & Swanson, 2013). In the US school system, as students transition from elementary to middle school, an increased emphasis on grades, test scores and academic tracking coincide with these developmental changes. The result of the poor fit between students' developmental needs and the context of middle schools is that, for many students, the transition to middle school is marked by declines in academic self-efficacy, motivation identity and engagement with school (see Eccles, Midgley, Wigfield, Buchanan, Reuman, Flanagan & MacIver,1993).

In all experiments, I used a methodological technique for maximizing both ecological validity and experimental control. As stimuli, students viewed a previously-recorded mathematics lesson on individual lap tops alongside their peers in their everyday classrooms. The lesson included cognitively demanding opportunities for higher order thinking intended to promote enduring conceptual understandings, such as comparisons across solution strategies including misconceptions. The pressure manipulation (stereotype threat; evaluative performance pressure; both or neither) was also delivered via video, which enabled within classroom condition assignment. Importantly, in order to distinguish between effects of pressure on learning versus performance, I varied whether the pressure manipulation was invoked before versus after the lesson and measured learning immediately following instruction as well as at a 1-week delay.

In Experiments 1 and 2, I examined impacts of these pressure sources separately. Experiment 1 tested impacts of stereotype threat. In Experiment 1a., I compared learning and engagement when stereotype threat was invoked either before learning or not at all. In Experiment 1b, I added a third experimental condition in which stereotype threat was invoked after instruction, in order to compare impacts of stereotype threat while learning with those of stereotype threat while testing. Experiment 2 tested effects of evaluative performance pressure during either learning or testing. In Experiment 3, I ran a 2 x 2 research design crossing stereotype threat and evaluative performance pressure to examine impacts of experiencing both sources of pressure simultaneously, as these pressures often co-occur in actual classrooms. In addition to examining overall impacts of these pressures, I considered student level factors that may moderate effects and support learning despite pressure. Together, these studies address the following overarching research questions:

1) **What sources of pressure help vs. harm learning?**

   I hypothesized that stereotype threat, due to its implications for identity would be especially likely to harm learning, while evaluative performance pressure would have greater potential to in some instances support learning through increasing motivating and effort.

2) **For which students does pressure help vs. harm learning?**

   In considering student level factors that might shape whether pressure helps vs. harms learning, I tested whether effects depended on either student's baseline EFs or gender, both of which have been found to moderate effects of pressure in testing context (e.g. Attali, Neeman & Schlosser, 2011; Levitt et. al, 2016; Maloney, Sattizahn, & Beilock, 2014; Maloney, Schaeffer& Beilock, 2013). I hypothesized that detrimental impacts of pressure would be greatest among students high in EFs, which has often been found to be the case when pressure is experienced while testing (e.g. Maloney, Sattizahn, & Beilock, 2014; Maloney, Schaeffer& Beilock, 2013). I hypothesized that beneficial effects of pressure could be larger among boys, as prior research has often shown larger performance advantages under pressure among males (e.g. Levitt et al., 2016; Attali, Neeman & Schlosser, 2011).

3) What student characteristics support learning despite pressure?

I examine whether students' learning orientations (motivations for engaging in academic behavior, Midgley et al., 2000) and academic efficacy (beliefs that one's academic efforts will result in desired outcomes, Midgley et al., 2000) can support learning despite pressure. Children's learning orientations and beliefs about their academic efficacy are well-documented predictors of academic achievement (e.g. Bandura, 1993; Blackwell, Trzesniewski &Dweck, 2007). I examine whether benefits of these student characteristics extend to high pressure contexts and explore whether these student characteristics might even protect against detrimental effects of pressure.

Theoretical Framing and Review of Relevant Literature

Classrooms can be stressful places for many students, with the pressures of children's larger socio-cultural contexts often taking shape in everyday interactions. In the post-recession educational landscape, characterized by increasingly rigorous content standards beginning in the earliest elementary grades (e.g., Bassok, Latham, & Rorem, 2016), heightened competitiveness of college admissions, and continued salience of high stakes assessments, many students feel a great deal of pressure at school (Luthar & Kumar, 2018; Wasserberg & Rottman, 2016). Coupled with this general evaluative pressure, children of color, who now represent the plurality of students in US public schools (NCES, 2017), may worry that their academic abilities will be judged based on negative stereotypes, a phenomenon often referred to as *stereotype threat* (see Steele & Aronson, 1995). Worries about being judged stereotypically can be heightened in the context of high stakes testing (Wasserberg & Rottman, 2016; Wasserberg, 2017), but are also experienced by many students during everyday classroom instruction (Larnell, Boston & Bragelman, 2014; Legette, 2018; Nasir et al., 2009, 2017).

These pressures can be especially salient in the mathematics classroom. Math is often a "high stakes" subject, meaning that students' scores on summative mathematics assessments are frequently used to evaluate teachers and schools in addition to the students themselves (e.g. Au, 2007; Croft, Roberts & Stenhouse, 2015; Dulude, Spillane, & Dumay, 2017). Along with these performance pressures, racial as well as gender stereotypes can be especially salient in math contexts (e.g., Chestnut, Lei, Leslie, & Cimpian, 2018; Bian, Leslie, & Cimpian, 2018; Miller, Eagly & Linn, 2015; Nasir & Shah, 2011). One reason that stereotypes can be especially influential in mathematics settings may be because mathematics ability is often viewed as something one either is or is not born with, depending not on effort but on innate talent or "brilliance" (Leslie, Cimpian, Meyer, & Freeland, 2015; Cimpian & Leslie, 2017).

Consequently, many students experience both stereotype threat (pressure stemming from increased salience of identity and resulting concerns about being judged through the lens of negative stereotypes) and evaluative performance pressure (pressure stemming from increased salience of evaluation) during every day mathematics learning opportunities. Stereotype threat and evaluative performance pressure share commonalities. Importantly, both sources of pressure can lead to intrusive thoughts and worries that interfere with cognition (Beilock & Carr, 2005; Schmader & Johns, 2003; Schmader, Johns & Forbes, 2008). However, the extent to which identity is explicitly implicated and potentially threatened differs. When experiencing stereotype threat, a primary concern is that oneself could be judged poorly, while in the case of evaluative pressure, the direct concern is about how one's performance will be judged. It is therefore important to distinguish between the roles of pressure to perform and stereotype threat as a specific type of pressure, as these sources of pressure may have different implications for children's mathematics learning and identities.

Pressure and Academic Achievement: Motivator or Threat?

How do these different but often overlapping pressures shape children's learning and academic engagement? Although this question has been considered from several disciplinary perspectives, the answer is far from clear.

On the one hand, pressure can act as a threat, resulting in distracting thoughts and worries. These intrusive thoughts impose upon executive functions (EFs), which are cognitive resources needed for attentional control, manipulation of mental representations, and task switching (Miyake et al., 2000; Schmader & Johns, 2003; Schmaeder & Beilock, 2012), making these valuable resources less available for engaging other academic tasks (Beilock & Carr, 2005; Schmader & Johns, 2003; Schmader, Johns & Forbers, 2008). An extensive research body demonstrates that experimental manipulations of either stereotype threat or evaluative performance pressure can harm academic performance (for reviews see Nguyen & Ryan, 2008- effects of stereotype threat or DeCaro et. al, 2011- effects of evaluative pressure). This work is reviewed more extensively in Chapter 2, but as one well-known example of stereotype threat harming performance, Steele & Aronson (1995) found that high achieving African American undergraduate students underperformed on a verbal test when the test was described as diagnostic of ability, or when participants were asked to report their race before the test. The largest performance decrements under both stereotype threat and pressure are often seen among high potential individuals, such the university students in Steele & Aronson's (1995) seminal study on stereotype threat, as well as individuals high in EFs more broadly, who otherwise have the highest performance (Beilock & Carr, 2005; Beilock & Decarro, 2007). This may be because these individuals are more likely to rely on cognitively demanding strategies to solve problems when not experiencing pressure (Maloney, Sattizahn, & Beilock, 2014; Maloney, Schaeffer& Beilock, 2013).

On the other hand, pressure can also act as a motivator, incentivizing increased effort and contributing to higher academic achievement (e.g. Angrist & Lavy, 2009; Leuven, Oosterbeek & Klaauw, 2010; Levitt, List & Sadoff, 2016). Indeed, a growing body of behavioral economics research indicates that raising the stakes of performance by introducing the possibility of earning incentives can result in improved academic performance (e.g. Angrist & Lavy, 2009; Leuven, Oosterbeek & Klaauw, 2010; Levitt, List & Sadoff, 2016; List, Livingston, & Neckermann, 2018). This work is reviewed more extensively in chapter three, but includes as one example a large randomized field experiment conducted with students in the Chicago area, in which Levitt, List, Neckerman & Sadoff (2016) found that raising the stakes of performance by introducing the possibility of earning incentives improved test performance among elementary school students. The prospect for increased pressure to lead to better performance is also supported by findings that GRE performance is higher under pressure (i.e. the real GRE) than in a low stakes situation (i.e. a voluntary experimental section of the GRE that participants could select to take following the actual exam) (Attali, Neeman & Schlosser, 2011). The performance advantage under high stakes settings or incentives is often found to be larger among males (Levitt et. al, 2016; Attali, Neeman & Schlosser, 2011). The greater performance boost with heightened pressure among males has been hypothesized as possibly stemming from gender differences in time sensitivity to rewards or to males putting forth less effort in low pressure settings (Levitt et. al, 2016; Attali, Neeman & Schlosser, 2011).

Much of the prior research on pressure and academic achievement has focused on adult’s performance in testing situations. However, children and adolescents also experience stereotype threat and pressure during everyday classroom learning opportunities (e.g. Larnell, Boston & Bragelman, 2014; Legette, 2018; Nasir et al., 2009, 2011, 2017; Nasir, Snyder, Shah & Ross, 2012). As compared with impacts of pressure during summative assessments, pressure experienced during day to day classroom instruction may have especially far-reaching consequences, shaping not only performance, but initial knowledge formation, academic identities and goals. Additionally, due to normative developmental changes across the lifespan, pressure may play a different role in shaping academic achievement among children and adolescents, as compared with impacts found in research conducted with adult populations.

Pressure and Academic Achievement during Early Adolescence: Developmental Considerations

In my dissertation, I examine the role of pressure in mathematics learning contexts during one particular developmental phase: early adolescence. Because adolescence is a pivotal time period for identity construction and for developing a sense of efficacy (see Erikson, 1959; Marcia, 1980; Swanson, Spencer & Petersen, 1998), experiences of either motivating or threatening pressure during this developmental period may set in motion recursive processes that have long range implications for young people's academic identities and aspirations.

Adolescence is characterized by rapid growth and development across physical, cognitive and affective domains (see Swanson, Spencer & Petersen, 1998). These rapid changes contribute to adolescence being a time of both heightened opportunity and risk, in which contextual supports and challenges can have far-reaching, recursive consequences for young people's life course outcomes. As described in Spencer & Swanson (2013, p. 19), "all humans have exposure to risks and protective factors, but the nature of the risks are different and the protective factors correlate with specifically experienced cultural supports and protective factors". Normative biological, cognitive and affective changes can increase all adolescents' vulnerability (Spencer & Swanson, 2013), and given the importance of identity formation during adolescence (see Erikson, 1959; Swanson, Spencer & Peterson, 1998) pressure experienced as a threat to identity may be especially significant for adolescents. Thus, during adolescence, "the invisible or unacknowledged fact and character of vulnerability" is especially pronounced (Spencer & Swanson, 2013, p 19).

As compared with either younger children or adults, adolescents' learning and motivation may be especially susceptible to external social pressures. Whereas characteristics of younger children's social and cognitive development are likely to reduce vulnerability to implicit social stressors such as stereotype threat and evaluative performance pressure, features of adolescent thought may heighten vulnerability to these same pressures. First, as compared with early adolescents, younger children are less likely to be aware of societal stereotypes, particularly those linking race and academic ability (McKown & Strambler, 2009; McKown & Weinstein, 2003). Additionally, the "egocentrism of early childhood" is characterized by a lack of awareness of other's thoughts such that young children are often not aware of or concerned with how others view them (see Elkind, 1967, Spencer 1985). This cognitive egocentrism can be protective against threats to identity and self- esteem, leading young children not to apply societal stereotypes to their sense of self as they might as they approach adolescence (see Harpalani, Qadafi & Spencer, 2013; Spencer, 1984, Spencer, 1985).

In contrast, characteristics of normative of adolescent development can heighten vulnerability to social pressures. For instance, the "egocentrism of adolescence" (see Elkind, 1967), which is characterized by a "failure to differentiate between the cognitive concerns of others and those of the self" (Elkind, 1967, p. 1025) can lead adolescents to believe that others are preoccupied with their behavior. A consequence is that, “in actual or impending social situations, the young person anticipates the reactions of other people to himself” (Elkind, 1967, p. 1030). Indeed, adolescence, and early adolescence in particular, represents a period of “extreme self-consciousness” and of “greater awareness of social expectations and inconsistencies” (Swanson, Spencer & Petersen, 1998, p. 21). These characteristics of adolescent development could heighten sensitivity to social pressures, including both evaluative performance pressure and stereotype threat. At the same time, like people of all ages, adolescents have exposure to a wide range of protective factors and supports that may for some students contribute to resilient learning outcomes even in the face of pressure (see Spencer, 2006, 2008).

Classrooms as Contexts for Development

Children’s learning and development occurs across multiple interconnected contexts that include the child’s immediate settings (for example, the classroom), as well as more distal contexts that nevertheless shape the child’s day to day experiences (Bronfenbrenner, 1979). Examples of more distal domains of influence include settings that the child does not encounter directly yet impact his or her everyday experiences (for example, policy decisions made in school district offices impacting classroom environments) as well as the broader socio-economic-political- cultural background, or macro system (for example, macro-economic patterns, societal beliefs and stereotypes). These multiple contexts of development interact with each other, and with a child’s own characteristics, to shape his/her learning and development, all within a particular socio-historical moment (Bronfenbrenner, 1979).

In the US and many other industrialized nations, classrooms are particularly influential settings for learning and development more broadly. From a purely time use perspective, children spend large portions of their waking hours in school settings. The average length of the school day for elementary aged children in the US is 6.8 hours, with most states requiring children to be in school for at least 180 days per year (NCES, 2018). Beyond the pure amount of time children spend in school, classrooms are often the spaces in which children first have meaningful interactions with adults beyond their immediate families and communities, encounter individuals of difference races and socioeconomic backgrounds, get feedback about their competencies relative to their peers and begin to engage in social comparisons.

The classroom spaces in which children spend such significant portions of their lives and navigate these developmental tasks are powerfully shaped by the broader cultural and sociopolitical context. As posited in ecological systems theory (Bronfenbrenner, 1979), ideologies, beliefs and pressures in the broader context shape children's experiences of their day to day environments (Bronfenbrenner, 1979). The role of these more distal, or macro-level, factors in shaping children's immediate contexts can be especially pronounced when considering children's classroom experiences. Importantly, the implications of these processes for children's learning and development are not evenly distributed; inequality within the broader society shapes children's immediate contexts of development (for discussion, see Spencer & Swanson, 2013; Velez & Spencer, 2018).

Stereotypes and high stakes testing policies are two ways in which the broader sociocultural context shapes children's learning experiences in uneven ways. For example, decisions about school funding, accountability structures and testing policies made at the state and federal level shape children's classroom experiences in ways that intersect with race, gender and social class, and often act to reinforce disadvantage and privilege (Au, 2009; Diamond & Spillane, 2002; Picower, B. & Mayorga, 2015). Additionally, school age children are aware of broadly held societal stereotypes, particularly along the lines of race (McKown & Strambler, 2009; McKown & Weinstein, 2003) and gender (Cvencek, Meltzoff &Greenwald, 2011), and these can influence children's interests and expectations for success in academically-relevant activities (e.g. Bian, Leslie & Cimpian, 2017). Together, these processes contribute to disparities in children's mathematics learning experiences, and contribute to mathematics achievement and participation remaining strongly patterned by student demographics, including race and gender (NSF, 2017).

Learning Despite Pressure: Protective Factors

Experiences of stress in the classroom are commonplace; nearly all students experience, at least at some point, potentially distracting worries and concerns while engaging in learning activities. However, despite the normative nature of at least some classroom stress, the type and intensity of stressors experienced, as well as the individual risk and protective factors students bring with them to an experience of stress, are not evenly distributed. Punitive high stakes testing pressures and concerns about being judged based on negative stereotypes in particular can intersect to create conditions of heightened challenge for children from marginalized groups (e.g. Wasserberg & Rottman 2016; Wasserberg, 2017).

How do these pressures shape children's learning and engagement in the mathematics classroom? While performance pressure and stereotypes can pose challenges to mathematics learning, its role in shaping children's learning outcomes is not deterministic. The Phenomenological Variant of Ecological Systems Theory (PVEST) is a theory of development that acknowledges the importance of both risk and protective factors in shaping children's iterative meaning making processes during an instance of stress engagement, within a particular context (Spencer, Dupree & Hartmann, 1997, Spencer et. al., 2006, Spencer, 2008). Thus, PVEST provides a helpful framework for disentangling pressure as distracting threat from pressure as motivating incentive. Whether a child exhibits positive learning outcomes despite pressure will depend on the child's *net vulnerability*, a function of both protective and risk factors (see Spencer, 2008), and the *net stress* of the experience itself, a function of the supports and challenges available in the context in which the pressure is encountered (see Spencer, 2008). Given the importance of meaning-making processes, it becomes clear that the role of pressure in shaping mathematics learning will not be deterministic or uniform. Instead, the content of the pressure itself, as well as children's risk and protective factors will shape children's meaning making, coping processes and learning outcomes during instances of stress engagement.

In my dissertation, I examine impacts of identity and non-identity threatening pressure sources experimentally elicited before a high quality yet challenging mathematics learning opportunity. I posit that whether a particular experience of pressure facilitates or interferes with learning will depend on how the pressure is experienced. Thus, in order to disentangle when pressure acts as motivating incentive vs. distracting threat requires taking into account the content of the pressure as this will have implications for children's meaning making and coping processes while engaging the pressure source.

Given the centrality of identity a particularly important dimension along which the content of pressure sources can vary is the extent to which identity is implicated. I predicted that pressure that implicates and potentially threatens a child's identity in math, as in the case of stereotype threat, would be especially likely to be experienced as threatening and to harm learning. On the other hand, I predicted that a pressure source that does not directly implicate a child's identity would have greater potential to in some cases be experienced as motivating and boost learning. My dissertation consists of a series of three experiments in which I first examine impacts of identity threatening and non-identity threatening separately and then together. This experimental progression is structured to help to distinguish between the roles of pressure to perform and stereotype threat as a specific type of pressure, as these sources of pressure may have different implications for children's mathematics learning and identities.

In addition to the content of the pressure source, particularly whether or not identity is implicated, many risk and protective factors contribute to an individual child's level of net vulnerability vs. resilience to heightened pressure during mathematics instruction. In my dissertation, I explore risk and protective factors that may contribute to some students being more or less vulnerable vs. resilient to these pressures. I examine in particular whether children's **baseline EFs**, **learning orientations** and **academic efficacy** act as risk or protective factors for learning in high pressure or threatening classroom contexts.

Children's views of themselves as students and beliefs about their academic efficacy are powerful predictors of learning generally (Bandura, 1993; Blackwell, Trzesniewski &Dweck, 2007; Eccles & Wigfield, 2002), and could be especially important in high pressure or threatening learning contexts. Students who have high academic efficacy and believe that their learning efforts will result in desired outcomes may be more likely to experience heightened pressure as motivator than as threat.

Likewise, students' goal orientations, or motivations for engaging in academic behavior, are important predictors of learning and achievement outcomes (e.g. Dweck, 1986; Cury et. al, 2002; Elliot et. al 2005), and could be particularly important in high pressure learning contexts. Goal orientations can be categorized into one of the following three classifications: Mastery Orientation (primary motivation for engaging in academic behavior is developing competence or understanding); Performance-Approach Orientation (primary motivation is demonstrating competence); or Performance-Avoid Orientation (primary motivation is to avoid demonstrating incompetence) (Midgley et. al, 2000). The mastery orientation has been associated with the most positive learning and achievement outcomes, while the performance-avoid orientation has been associated with negative outcomes (Midgley et. al, 2000). Possessing a mastery learning orientation, could be protective against impacts of pressure, as children who are motivated to learn for learning's own sake might be less impacted by external pressures.

Finally, because a mechanism through which pressure can interfere with or facilitate learning is its impacts on children's cognitive engagement during instruction, children's baseline cognitive resources, EFs, may contribute to whether pressure harms or helps learning. I test whether having high baseline EFs acts as a risk or protective factor for coping with identity threatening and non-identity threatening pressure during conceptual math instruction.

Race, Gender and Mathematics

Relationships between race, gender and mathematics achievement are complex. Although no inherent racial differences in academic ability exist, racialized patterns of achievement persist within the US; these patterns are often especially pronounced in mathematics test scores (NAEP, 2009; NAEP, 2011; NAEP, 2015; NCES, 2016). Mathematics racial achievement gaps are in part shaped by uneven learning opportunities, as students of color are more likely to have math teachers who are inexperienced or not qualified to teach math and to consequently receive low quality math instruction (Rahman et al., 2017), and may be further compounded by experiences of racial microaggressions and stereotype threat in the classroom (Goings & Bianco, 2016; Larnell, Boston & Bragelman, 2014; Legette, 2018; Nasir et al., 2009, 2017). Concerns about being judged stereotypically can lead to intrusive thoughts and worries that tax executive functions (EFs), which could interfere with children's engagement with instruction, reducing learning, even when high quality instruction is available. Thus, experiences of stereotype threat in the classroom may further compound the gap in opportunity stemming from unequal access to high quality instruction. Similarly, although males and females do not differ fundamentally in mathematics aptitude, gender gaps in mathematics achievement and participation persist, with boys continuing to exhibit higher performance on standardized tests, particularly at the highest levels of achievement (Hyde et al., 2008; Reardon et al., 2018).

Additionally, gender and race intersect in complex ways to shape children's mathematics learning experiences. In mathematics and science contexts, African American and Latinx female students can experience a double stereotype threat (e.g. Brown & Leaper, 2010; McGee &Bentley, 2017; Young, Young & Capraro, 2017), yet often are able to draw on supports to navigate threatening mathematics learning contexts, although this success is not without strain (McGee & Bentley, 2017; McGee & Spencer, 2012). Indeed, African American and Latinx females often exhibit high mathematics achievement relative to male students of their same race or ethnicity. As one example, the male advantage in mathematics test performance described above is driven nearly entirely driven by White and Asian students (Reardon et al., 2018).

Although males of color may not face a double threat (Brown & Leaper, 2010) in mathematics contexts, negative racialized stereotypes pertaining to males can be especially strong, and become increasingly so as children approach adolescence (Ellis, Rowley, Nellum & Smith, 2018; Goings & Bianco, 2016). Thus, stereotypes, along with many other contextual factors, can pose unique challenges to academic success for African American and Latinx male students (Ellis et al., 2018; Swanson, Cunningham & Spencer, 2005). These include as one important example a near complete absence of male teachers and male teachers of color in particular in elementary classroom settings (e.g. Goings & Bianco, 2018). Given consistent gender differences in academic achievement among students of color (e.g., Kaba, 2005;

Saunders, Davis, Williams, & Williams, 2004; Reardon et al., 2018), it is important to consider the possibility of gender variation in race related experiences, such as stereotype threat (for discussion, see Chavous et al., 2008).

Mathematics during early adolescence

Mathematics during the early adolescent years can be a source of both challenge and opportunity. Cognitive changes occurring during adolescence increase young people's capacity for complex thought and social comparisons- but also heighten vulnerability to negative feedback from the environment. In the US school system, as students transition from elementary to middle school, an increased emphasis on grades, test scores and academic tracking coincide with these developmental changes. The result of the poor fit between student's developmental needs and the context of middle schools is that, for many students, the transition to middle school is marked by declines in academic self-efficacy, motivation identity and engagement with school (Eccles et. al, 1993; Eccles, 2004; Wigfield & Eccles, 2002). These can be especially pronounced in math. Math grades, intrinsic motivation and interest often decline during the transition from elementary to middle school (Fredericks & Eccles, 2002; Gottfried, Fleming, & Gottfried, 2001; Gottfried, A.E. Marcoulides, Gottfried, G. A., Oliver & Guerin, 2007; Wigfield, Eccles, Schiefele, Roeser, & Davis-Kean, 2006).

During the upper elementary and middle school years, the math curriculum becomes increasingly abstract and conceptual. For example, it is during this time that students must make the conceptual leap from whole numbers to fractions (National Mathematics advisory, 2008). This lays the groundwork for engaging successfully in abstract quantitative reasoning, but moving from thinking in terms of whole numbers to ratios is a major source of difficulty for many students (Siegler et al., 2012; Hecht, Close & Santisi, 2003).

Engaging with abstract math concepts within the context of high quality math instruction provides early adolescents the opportunity to exercise their growing capacities for abstract thinking and reasoning (Blumenfield, Kempler & Krajci, 2008; Middleton & Midgley, 2002), and students' successful engagement with conceptual math during middle school is predictive of positive life course outcomes, including enrollment in advanced math courses during high school and college going (Spielhagen, 2006). Indeed, access to high quality, conceptual math instruction within the context of a supportive instructional environment during middle school can powerfully support not only students' math achievement, but also academic identity and achievement more broadly (Wynne & Moses, 2008; Grant, Crompton & Ford, 2015). While grappling with challenging and abstract math concepts during middle school can act as a leverage point for increased interest and success in advanced math and careers, the increasingly abstract nature of the mathematics curriculum can also be a source of challenge.

In addition to challenges stemming from the nature of the curriculum itself, pressure and anxiety or, alternatively, disengagement, can both create challenges to math success during the middle school years. On the one hand, as math performance is often high stakes, for example being used in course placement and grade promotion decisions, students may feel so much pressure to perform well that anxious ideation interferes with their learning and performance. On the other hand, if students do not see mathematics as having personal relevance within their current or future lives, students may feel little motivation to perform well and may disengage during math learning and testing opportunities. Thus, both caring too much (i.e. anxiety) and caring too little (i.e. disengagement) can pose challenges to mathematics success during middle school.

Contributions of the Current Work

As discussed above, the broader sociocultural context shapes children's mathematics learning experiences. One important consequence of this is differences in the extent to which pressure imbues children's classroom learning experiences, the types of pressures present and how these pressures are experienced. Evaluative performance pressure and stereotype threat have been well studied in performance contexts. However, this work has not fully taken learning and instruction into account- leaving open the question of how these pressures may impact initial learning and could differentially incentivize or impair learning for students in a systematic way.

Experiences of pressure during everyday instruction could have especially far-reaching consequences, particularly in a field such as mathematics where learning is cumulative. While experiencing pressure during testing can harm performance and possibly future motivation and learning, it does not directly affect a student's knowledge base beyond any benefits of successfully retrieving test items during a test, such as those demonstrated in the test effect literature (see Roediger & Karpicke, 2006). Effects of pressure during instruction conversely are likely cumulative as students progressively build new understandings on a foundation of prior knowledge. If experiences of pressure prevent students from fully benefitting from instruction at one time point, subsequent learning opportunities could also be affected. In my dissertation, I build upon understandings gleaned from the study of pressure in performance contexts to begin to build our understanding of the role of these pressures in shaping children's initial learning.

Based on the damaging effects of pressure on available cognitive resources and test performance (e.g. Schmader & Johns, 2003; Schmaeder & Beilock, 2012), it is likely that these stressors may also impact learning, as learning is a cognitively demanding endeavor- perhaps even more so than testing. Learning requires that students are able to attend to instruction, hold in mind related prior knowledge, and incorporate new understandings gleaned from instruction into existing knowledge schemas. If cognitive resources are compromised due to stereotype threat or incentivized performance pressure, students' ability to successfully engage in this cognitively demanding process of learning may be compromised. At the same time, because mathematics is a subject in which students often exhibit disengagement (Sullivan, Tobias & McDonough, 2006; Martin et al, 2012), it is also possible that increased pressure could boost student learning and achievement through incentivizing increased focus and effort.

I focus on the impact of pressure within a particular instructional context, that of a cognitively demanding conceptual mathematics lesson. The lesson's objective was to teach students to compare ratios with different denominators. It centered on a comparison between a correct strategy (lowest common multiple) and a misconception (subtraction). The act of comparing and contrasting solutions, including misconceptions, is a recommended educational practice that has been shown to promote deep learning (Durkin & Rittle-Johnson, 2012; Richland & McDonough, 2010; Rittle-Johnson & Star, 2007). At the same time, comparing correct and incorrect solution strategies places high demands on student's cognitive resources, as students must be able to learn from the misconception discussion, but ultimately inhibit it as a prepotent response, and encode the correct strategy (Zaitchik, Iqbal, & Carey, 2014). Just as performance on challenging test items is most harmed during experiences of stereotype threat or performance pressure, this type of high quality- high demand learning could be especially compromised by situational stressors that tax cognitive resources. In order to maximize ecological validity while also allowing for controlled stimuli (i.e. ensuring that all students are exposed to the exact same high-quality instruction), I implement a research methodology in which students view a previously-recorded conceptual math lesson alongside their peers in their normal math classrooms.

I focus on the role of pressure in shaping mathematics learning and interest as children transition from middle childhood to early adolescence. This is a pivotal time period for identity construction (Erikson, 1968), and children's learning experiences during this time may have especially important implications for their academic trajectories.

Overview of Studies

In a series of three classroom based experiments, my dissertation examines how experiencing increased pressure during cognitively demanding mathematics instruction impacts 5th grade students' mathematics learning and interest. I focus on two different but often overlapping sources of pressure that many students experience during everyday math instruction: stereotype threat (pressure from increased salience of a self-relevant negative stereotype, see Steele & Aronson, 1995) and incentivized performance pressure (pressure from the possibility of obtaining or losing an incentive). Importantly, these pressure sources differ in the extent to which identity is implicated. Thus, I hypothesized that while stereotype threat would most likely have harmful effects on mathematics learning and engagement, incentivized performance pressure could in some cases have the potential to boost learning and engagement through increased motivation and effort.

In the first portion of my dissertation, I examine these pressure sources separately. Chapter 2 tests effects of stereotype threat. Experiment 1a tests impacts of an implicit race based stereotype threat prior to instruction on mathematics learning and interest among students of color. Mathematics learning and interest is compared when stereotype threat is invoked either before instruction or not at all. Experiment 1b tests effects of stereotype threat prior to instruction and stereotype threat prior to testing, comparing learning and interest when stereotype threat is invoked either before instruction, before testing or at neither time. Chapter 3 examines effects of incentivized performance pressure prior to instruction or testing. Mathematics learning and interest is compared when incentivized performance pressure is invoked before instruction, before testing or not at all. In this portion of my dissertation, I consider the role of student gender and baseline executive functions (EFs) in shaping impacts of stereotype threat and pressure. Finally, as stereotype threat and incentives often co-occur in actual classrooms, in the second portion of my dissertation, I examine the role of both pressure sources in shaping children's mathematics learning and interest. This experimental progression is intended to contribute understanding of similarities and differences between stereotype threat and evaluative performance pressure, helping to better distinguish between pressure to perform and stereotype threat as a particular type of pressure in order to disentangle when pressure acts as a distracting threat vs. motivating incentive.

Specifically, in Experiment 3, I ran a 2 x 2 research design crossing stereotype threat and incentivized performance pressure before learning. Participants were randomly assigned to one of four design cells that resulted from crossing two between-subjects factors: stereotype threat (vs. not) and incentivized performance pressure (vs. not). In Chapter 4, I examine impacts of stereotype threat and incentivized performance pressure, and, as in the earlier chapters, consider the role of gender and baseline EFs. In this experiment, I also gathered additional measures of student characteristics, including learning orientations, academic efficacy, novelty preferences and mathematics anxiety. In Chapter 5, I explore student characteristics that can promote learning, even in high pressure or threatening contexts. Specifically, I examine relations between students' learning orientations, academic efficacy, attitudes towards novelty and mathematics anxiety and learning gains during the lesson. I test whether these relations differ or are similar in high and low-pressure contexts. The goal of these analyses is to identify student characteristics that support learning generally, and to examine whether these relationships extend to high pressure contexts.

CHAPTER TWO: STEREOTYPE THREAT DURING LEARNING AND TESTING

Introduction

Stereotype threat— a situational context in which individuals are concerned about confirming a negative stereotype — is often shown to impact test performance, with one hypothesized mechanism being that cognitive resources are temporarily co-opted by intrusive thoughts and worries, leading individuals to underperform despite high content knowledge and ability (see Schmader & Beilock, 2012). A large body of research indicates that experiences of stereotype threat in performance contexts can lead capable individuals to underperform relative to their knowledge base (for reviews, see Nguyen & Ryan, 2008; Pennington et al., 2016). The most pronounced performance decrements under stereotype threat are often seen among individuals with the greatest potential to excel, such as those high in EFs (see Beilock, 2008; Maloney, Schaeffer & Beilock, 2013).

This chapter examines whether stereotype threat may also impact initial student learning and knowledge formation when experienced prior to instruction (Experiment 1a), and compares impacts of stereotype threat during learning with those of stereotype threat during testing (Experiment 1b). Specifically, in Experiment 1a., a race based stereotype threat was invoked implicitly either before instruction or not at all. Experiment 1b replicates findings from Experiment 1a and also compares effects of stereotype threat while learning with those of stereotype threat while testing. In Experiment 1b, stereotype threat was invoked either before instruction, before testing, or not at all.

Stereotype Threat and Performance

Much of what we know about stereotype threat comes from studies conducted in performance contexts- in which participants are asked to demonstrate their knowledge of previously-learned material. In the first study on this phenomenon, Steele & Aronson (1995) found that high-achieving Black undergraduate students underperformed on a verbal test when the test was described as diagnostic of ability, or when participants were asked to report their race before the test.

Since this seminal study, effects of stereotype threat on performance have been studied extensively. In these studies, stereotype threat is most often invoked by giving participants a prompt before a test that includes some combination of increasing identity salience, such as by asking participants to report their race or presenting the study as being concerned with how individual differences influence performance, and raising the stakes of performance, such as presenting the task as diagnostic of ability. For example, in Steele & Aronson's (1995) seminal study, stereotype threat was invoked among African American adult participants by describing the study as "being concerned with 'various personal factors involved in performance' and as 'diagnostic' of participants' ability. In contrast, in the control condition, the test was described as simply aimed at better understanding factors involved in problem solving. African American participants who received the former instructions underperformed relative to African American participants who received the control prompt and relative to white participants who received either prompt.

Along with additional research on effects of racial stereotypes on academic performance among African Americans (for review, see Nguyen et al., 2008), researchers have extended the study of stereotype threat's effects on performance to several different populations about which negative stereotypes exist, including: females (math/spatial ability: Shapiro, 2012; Smith, 2007); the elderly (memory: Chasteen, 2005; Hess, 2003), and low SES individuals (intellectual ability: Croizet & Claire, 1998). Experimental invocations of stereotype threat have been shown to lead to performance decrements even among individuals who are members of groups about which there is not a widely-shared negative stereotype, such as white men when told that their math performance will be compared with that of Asian men (Aronson, Lustina, Good, & Keough, 1999).

However, notwithstanding the shared capacity for vulnerability to stereotype threat, the burden of actually contending with stereotype threat falls disproportionately on individuals whose identities are negatively stereotyped. In US K-12 classrooms, African American and Latinx students shoulder the greatest burden of stereotype threat. Alongside a student population that is more diverse than ever before (NCES, 2017), negative stereotypes about the academic abilities of people of color persist (Nasir, 2017), and students are aware of these stereotypes from a young age, with most children being aware of broadly held stereotypes about the academic abilities of people of color by at least age ten (McKown & Weinstein, 2003; McKown & Strambler, 2009). Stereotype threat, therefore, is a source of stress that disproportionately affects students of color and may contribute to achievement gaps.

Although the majority of research on stereotype threat has been conducted with adults, a growing body of research documents that stereotype threat can also harm children's test performance. Much of this research has examined effects of stereotype threat on girls' performance on math, science and spatial skills tests (for meta-analytic review, see Flores & Wicherts). Yet, gender differences in children's mathematics performance are quite small and in many cases nonexistent at the K-12 level (see Reardon, 2018), suggesting that racial stereotypes may be more salient and thus play a greater role in shaping students' experiences in the context of K-12 classrooms. A few notable studies have shown effects of racial stereotypes on children's academic performance. For example, McKown & Weinstein (2003) found that when a test was described as diagnostic of ability, 6-10-year-old African American and Latinx students who were aware of racial stereotypes about academic ability underperformed. Attending a predominantly minority school does not appear to inoculate students to detrimental effects of stereotype threat. Indeed, Wasserberg (2014) found stereotype threat effects among African American and Latinx 3rd to 5th graders attending highly segregated urban schools (>95% students of color) who were stereotype aware. In this study, when racial group membership was made salient by asking students to indicate their race prior to testing, students who were aware of academic stereotypes showed decreased performance under diagnostic testing conditions. Relative to the younger students, 5th graders were more likely to be stereotype aware and thus more impacted by stereotype threat (Wasserberg, 2014).

Stereotype Threat and Cognition

More recently, research on stereotype threat in testing contexts has moved beyond documenting effects on performance, towards understanding mechanisms and moderators. While many different mechanisms and moderators have been explored (for review, see Pennington et al., 2016), much of this work highlights that cognitive resources play a central role as both moderators and mechanisms of stereotype threat's effects. In a review article, Rydell & Boucher (2017) underscore the central role of cognitive recourses. Under stereotype threat, the discrepancy between one's positive view of self and the negative stereotype about one's in-group creates an unsettling conflict. Efforts to resolve this tension (e.g. distancing self from the stereotyped group, suppressing negative thoughts) can tax cognitive resources, leaving individuals under threat with fewer cognitive resources available to devote to grappling with test items, and this reduction in cognitive resources is the proximal cause of underperformance in the face of threat (Rydell & Boucher, 2017).

Stereotype threat's impacts on performance depend on both test-taker's baseline cognitive resources and how cognitively demanding test items are. Performance under stereotype threat is most affected on challenging, cognitively demanding problems, such as those that require comparison or suppression of a prepotent response (e.g. Davies, Conner, Sedikides, & Hutter, 2016; Maloney, Sattizahn & Beilock, 2014). Additionally, performance effects are often largest among individuals who are high in baseline EFs, as these individuals are more likely to rely on cognitively demanding strategies to solve problems when not experiencing stereotype threat (Beilock & Carr, 2005; Beilock & Decarro, 2007; Maloney, Sattizahn, & Beilock, 2014; Maloney, Schaeffer, & Beilock, 2013). Research that attends to cognition, taking into account both the cognitive demand of tasks and resources of individuals, can help to elucidate stereotype threat effects.

Stereotype Threat and Learning

Learning is a cognitively demanding process, in many ways more imposing on attentional resources than test-taking. Students must attend to instruction, hold in mind related prior knowledge, and incorporate new understandings gleaned from instruction into existing knowledge schemas. Stereotype threat's effects on cognitive resources make it likely to disrupt learning, particularly during high quality, cognitively demanding instruction as students experiencing stereotype threat may be less able to engage in opportunities for higher order thinking that promote enduring conceptual change. Despite this concerning possibility, our understanding of how stereotype threat impacts children's learning remains limited, as most research on stereotype threat has been conducted in testing contexts. While research on the effects of stereotype threat on initial learning pales in comparison to the extensive literature on stereotype threat in testing contexts, the relatively few studies that have focused on stereotype threat during initial learning opportunities indicate that stereotype threat's effects can indeed extend to initial learning (see Rydell & Boucher, 2017 for review).

The majority of research on stereotype threat effects on learning has focused on gender stereotypes about women's math and spatial abilities. In a series of studies, Rydell, Rydell & Boucher (2010) found that invoking negative stereotypes related to women's math and spatial ability harmed adult women in: learning to use a novel type of math involving using symbols to determine which equation to insert values into (Study 1) and in learning to solve modular arithmetic problems (Study 2). Stereotype threat has also been shown to harm adult women in learning to complete visual search efficiently (Rydell, Shiffrin, Boucher, Van Loo, & Rydell, 2010). Additionally, Mangels et al. (2012) found that experiencing stereotype threat decreased the extent to which adult women learned from mistakes they made during a challenging math test. In this study, the researchers used brain imaging and behavioral measures to compare women's responses to, and learning from, accuracy feedback during a GRE-like exam, under threatening and non-threatening conditions. They found that for participants experiencing stereotype threat, increased emotional salience of negative accuracy feedback (as evidenced by brain imaging) predicted less engagement with an optional tutorial showing how to solve the incorrect problem and worse learning from the tutorial even when participants did engage with it (Mangels et al. 2012). This pattern of results suggests two pathways whereby stereotype threat may harm learning: decreased engagement leading to less effort and cognitive load interfering with attempts to learn.

Although stereotype threat's performance effects were first documented among African American participants in response to negative racial stereotypes, even less research has examined impacts of stereotype threat on African American students' learning. In one of the few studies to do so, Taylor & Walton (2011) found that experiencing stereotype threat while learning esoteric vocabulary words decreased learning among African American adults.

While the findings reviewed above indicate stereotype threat can impact initial learning, most of this work has tested effects of stereotype threat among adults when asked to learn circumscribed tasks in laboratory contexts, limiting the ecological relevance of this work and its applicability to educational contexts. Additionally, the instructional content were all relatively low demand (for example, asking participants to memorize new definitions or problem-solving procedures) and covered topics outside standard curricula. Additionally, the normal bustle and diversions of day to day classroom contexts were absent from these tightly controlled laboratory studies. Finally, while much of this work has examined stereotype threat's effects on learning in among adults, learning has especial relevance and importance for children.

In a 2012 article in *Educational Psychology Review*, Appel & Kronberger underscored the need for more research on stereotype threat in learning contexts, calling for greater participation from educational psychologists in this endeavor. However, very few studies on stereotype threat during learning opportunities have been conducted since this time- and even fewer by educational psychologists in ecologically relevant contexts. With the exception of two studies that examined differential responses to performance feedback in threatening and non-threatening conditions and that considered the implications of these different responses for learning (Forbes, Duran, Leitner & Magerman, 2015; Mangels et al., 2012), every study included in Rydell & Boucher's (2017) review chapter on stereotype threat and learning was conducted prior to 2012. Additionally, despite increased participation from educational and developmental psychologists in research on stereotype threat in performance contexts, very little research has examined effects of stereotype threat on children's learning.

The two experiments reported in this chapter extend the study of stereotype threat to consider impacts on children's classroom learning during a high quality conceptual math lesson Both studies examine stereotype threat effects among early adolescents in the context of mathematics instruction that requires higher order thinking, which is ideally the case in everyday high-quality math instruction. Specifically, the lesson covered ratio, which is a topic within the normal math curriculum for children this age (National Governors Association Center for Best Practices, 2010). As described more fully in the methods section, the lesson prompts students to compare and contrast correct and incorrect strategies for comparing ratios. Mathematics instruction that incorporates "desirable difficulties" (Bjork, 1994), such as comparing and contrasting multiple solution strategies including misconceptions, can powerfully promote enduring conceptual understanding and reduce misconception endorsement (Begolli & Richland, 2016; Richland & McDonough, 2010; Rittle-Johnson & Star, 2007). However, this type of instruction places high demands on learners' EFs, as students must hold in mind and draw connections between multiple solution strategies, while also inhibiting the pre-potent misconception (Zaitchik, Iqbal, & Carey, 2014). If students are experiencing cognitive load due to stereotype threat, they may be less likely to realize the full benefits of this type of instruction.

Experiment 1a

Methods

Participants

All 5^{th} grade students at two K-6 charter schools were invited to participate in year 1, yielding 118 participants (3 classes at school 1; 2 classes at school 2). An additional cohort of students from school 1 was invited to participate in year 2, yielding an additional 51 students (3 classes). Thirty-three students who were absent on one or more study days and did not have complete data were excluded from analysis. Nearly all participants were members of minority groups that are negatively stereotyped academically. The majority of students (127) self-reported their race/ethnicity to be Black/African American. Additional students identified as Hispanic/Latino (2), or both Hispanic/Latino and Black/African American (6). No students identified as Asian and only one student, who was excluded from further analysis, identified as White. The remaining 135 participants were nearly evenly split by gender with slightly more girls (Control condition: 33 girls, 31 boys; Stereotype Threat (ST) condition: 37 girls, 34 boys). The distribution of participants' race/ ethnicity identification was comparable across conditions.

Teacher and Experimenter Demographics

Children's regular math teachers were present in the classroom throughout the study. In all but one classroom, the teacher was a white female. In one class at school 1, the teacher was a Black/African American male. The primary experimenter for all classrooms was a white female.

Design and Procedures

Procedures were administered during three visits over a two-week period at each school using a pretest, lesson and immediate posttest, delayed posttest design. Participants were randomly assigned within each classroom to either the Learning under a Stereotype Threat (ST) condition or Control condition to minimize variability across schools and teachers. Random assignment of participants to conditions within classrooms ensured that there were no differences in ethnic backgrounds of experimenters or teachers between experimental groups.

- At Visit 1, students completed a group-administered pretest to assess their starting understanding of ratio. This assessment, as well as the immediate and delayed posttests, measured conceptual and procedural understanding of ratio as well as misconception usage- the frequency with which students incorrectly made use of subtraction to attempt to solve problems requiring use of ratio concept (for test properties, see Begolli & Richland, 2016). An example assessment is available in the Appendix.

- At Visit 2, students interacted individually with a videotaped lesson on ratio, which centered on a comparison between two common solution strategies for solving ratio problems – one a frequent misconception (Subtraction) and one a correct solution (Least Common Multiple, or LCM). Greater details on the lesson and math assessments are provided below.

Stereotype threat was manipulated prior to the video lesson, before learning. Specifically, at the start of the video lesson, students in the ST condition viewed a video screen in which the following prompt was visible and read aloud: *"Thank you for being part of this study. Please turn to page 2 in your packet and fill in your race. Professors at the University of Chicago are very interested in knowing about you and how you perform on the test after the lesson. Your information is very important to us, because we want to learn how to best help kids like you learn math."* Students in the Control condition viewed a video screen in which the following prompt was visible and read aloud: *"Thank you for being part of this activity. Please turn to page 2 in your packet and write today's date. We're glad that you are doing this activity because it will help us learn about some of the best ways to teach kids math."* In both conditions, the prompt was visible on the screen for 30 seconds. With the exception of this introduction, the video was otherwise identical between the two conditions.

This stereotype threat manipulation is quite similar to other manipulations of stereotype threat- particularly those employed with children in school contexts (e.g. Wasserberg, 2014; McKown & Strambler, 2009). This manipulation of stereotype threat involved both increasing identity salience (by asking participants to provide their race and describing the study as being focused on better understanding how kids like you learn math), and raising the stakes (by informing students that an important aim of the study was to see how well they performed on a test after the lesson).

All students completed a **mathematics posttest** immediately following the video lesson. Following the posttest a subset of the students also completed measures to gather data on their situational interest and mathematics engagement (more detail on these measures are provided below). An abbreviated version of the **Situational Interest Survey** (Chen, Darst, & Pangriza, 2001) was administered to a subset of participants in both conditions, with test reliability leaving two measures: instant enjoyment and exploration intention. A subset of students in both conditions were also given the option to work on non-required **math puzzles** during free time after finishing the intervention video and posttest.

- At Visit 3, students completed a delayed posttest and a measure of EF, the d2 test of attention, (Brickenkamp & Zillmer, 1998), followed by a short demographics survey..

Materials

Video instruction. All participants watched an instructional video introducing the concept of ratio and strategies for comparing ratios. The instructional video was originally recorded as a live, semi-scripted lesson on ratio by a teacher with a diverse class of 5th grade students recruited for the recording. The teacher in the video was a white female and the students visible in the lesson frame were recruited to include both boys and girls of African American, Hispanic, Asian, and Caucasian race/ ethnicity.   The format of the lesson was an interactive class discussion. To simulate the active participation afforded by real classroom discussions, the video prompted students who were watching the video on a computer at key points during the lesson to answer questions on a worksheet paper, often the same ones posed to the students in the video, asking them to make inferences and draw connections with the lesson content. The lesson centered on a comparison between two common solution strategies for solving ratio problems – one a frequent misconception (Subtraction) and one a correct solution (Least Common Multiple, or LCM).

Mathematics assessments. At all three visits, students completed a mathematics assessment measuring both their conceptual understanding of ratio and their ability to solve problems that require ratio comparisons (for test properties see Begolli & Richland, 2016). The tests assessed students' ability to produce correct procedures to solve ratio problems like the one shown in the video lesson, as well as their ability to adapt procedures to novel problem types and contexts and to explain their reasoning for using a given solution strategy. The assessments also allowed for measurement of misconception usage- the frequency with which students incorrectly made use of subtraction to attempt to solve problems requiring use of ratio concept. The same test items were used on the assessments at each visit, but question order varied. Repeating assessment items allowed me to most closely assess changes in performance between sessions. No feedback was provided after each test, so it is unlikely that student learning would improve based on repeated tests alone. All assessments contained a mix of multiple choice and free response questions. An example assessment is available in the appendix. All free response questions were coded for accuracy, and all work shown was coded for validity of the attempted strategies, by at least two coders who were blind to participant condition (with Cohen's Kappa's between 0.88 and 1.00). The coding manual is available in the appendix.

Situational Interest. A subset of the students also completed an abbreviated version of the Situational Interest Survey (Chen, Darst, & Pangriza, 2001), a 20-item instrument designed to measure five components of situational interest (instant enjoyment, attention demand, novelty, challenge and exploration intention) in middle school students. I chose to shorten this previously validated scale due to time constraints associated with conducting research in school settings. In the abbreviated 10-item version, students rated from 1 to 5 the extent to which they agreed or disagreed with two items measuring each of the five components of situational interest. As discussed in Gogol et al. (2014), because testing time in educational research is typically scarce, the use of long scales to assess motivational-affective constructs can be problematic and short scales and even single-item measures can often be as useful as their lengthier counterparts for assessing motivational and affective constructs including academic anxiety, academic self-concept (Gogol et al, 2014) and Math Anxiety (Nunez-Pena, Guilera, Suarez-Pellicioni, 2014).

In order to assess the reliability of the shortened assessment, Cronbach's alphas for each of the original constructs of situational interest were calculated: instant enjoyment 0.778; exploration intention 0.674; Novelty 0.573; Attention demand, 0.572; and Challenge 0.477. Subscales with Cronbach's Alphas < 0.65 were excluded from further analyses, leaving two constructs of situational interest: instant enjoyment and exploration intention.

Math puzzles. Once they had completed the immediate posttest, a subset of students was given the option to work on four math puzzles. Students were informed that these puzzles were *optional* and they could choose whether or not to complete them. The math puzzles did not test specific content, but instead involved detecting patterns and completing sequences - a type of math task often considered fun and engaging by students. The students' other option was to independently read quietly. We recorded the number of puzzles each student attempted.

Executive Function (EF). EF was assessed using the d2 Test of Attention, a measure of sustained and selective attention and inhibitory control, normed with US and German children, adolescents, and adults (Brickenkamp & Zillmer, 1998). The task requires participants to search for target characters ("d"s with two dashes surrounding it) from among perceptually similar distractors (e.g., "d"s with one dash, "p"s with two dashes) under a time pressure. The focal outcome score analyzed in this study was the total number of items processed minus errors (TN-E), which yielded a range of 188-665. Internal consistency for this measure is high across the literature ($\alpha \geq 0.8$) (Clark, 2005). Test-retest reliability is also high at one day ($\alpha > 0.8$), but declines over time (Clark, 2005). Additionally, TN-E score correlates with other measures of attention and EF, including Stroop and Tower of London, supporting the validity of this measure (Clark, 2005). The task was group-administered to each class.

Results

Sample

Results for 135 students for whom I obtained complete data are included in the analyses reported below. Students at school 2 and at school 1 in year 2 ($n = 79$; 43 ST) also were administered the situational interest survey, and students at school 2 ($n = 40$; 21 ST) were administered the optional puzzles.

Analytic Plan

I first report an analysis of pretest performance to ensure that random assignment was successful, and then describe analyses pertaining to my main questions of interest: how student EF and stereotype threat during instruction impacted learning. I examine this using three types of outcome measures. These measures were used to examine main effects on learning, as well as to test the hypothesis that high EF students would be most affected by the threat manipulation (with high EF students defined as those scoring above the median on the d2).

- First, I examine immediate and sustained uptake of the misconception (subtraction) presented in the lesson, calculating uptake of the misconception immediately (proportion of problems solved with the misconception strategy at immediate posttest minus misconception use at pretest) and sustained over a delay (proportion of problems solved with the misconception strategy at delayed posttest minus at pretest). Any gains in use of the misconception after the lesson may indicate that the student memorized the first solution that was presented, without engaging in the higher order comparative processing within the lesson, which would have made clear that this is an invalid way to solve the problem. Thus, misconception gains could signal that students are motivated to memorize everything presented in the lesson in order to be able to perform well - at the possible expense of the type of deep, critical engagement with comparative instruction that promotes enduring conceptual understanding. Additionally, if EF resources were compromised, these participants may have been unable to engage in that higher-level comparative thinking.

- Second, I examine learning as defined as immediate and sustained gains in accuracy. Accuracy was calculated as correct set up and reasoning on a combination of procedural and conceptual items. Immediate learning was operationalized as accuracy gains from pretest to posttest, and sustained accuracy was operationalized as gains from pretest to the posttest administered a week after instruction.
- Lastly, I analyzed the impact of experiencing stereotype threat during instruction on the extent to which students reported that they enjoyed the lesson and desired to learn more about the topic covered in the lesson, as well as their likelihood of actually choosing to engage in a non-required math activity.

Performance at Pretest

Neither frequency of misconception use nor overall accuracy differed by condition at pretest (both ps > .5) or between high and low EF students (both ps > .4). Mean misconception use and accuracy at pretest for students in the two learning conditions are reported in Table 2.1.

Effects of Stereotype Threat Manipulation Overall

In a series of regressions, EF (entered as a continuous variable with values ranging from 136 to 665) and Condition assignment (binary coded: 0 = Control; 1 = ST) were first used to predict immediate and sustained changes in misconception use and content knowledge in the overall sample. Mean immediate and sustained gains in misconception use and accuracy for students in the Control and ST learning conditions are shown in Table 2.1, and all results and effect sizes from the regression are available in Table 2.2.

	Pretest Means (SE)		Immediate Gains (SE)		Sustained Gains (SE)	
	Misconception	Accuracy	Misconception	Accuracy	Misconception	Accuracy
Control *n = 64*	0.46 (.06)	0.15 (.02)	-0.11 (.04)	0.18 (.03)	-0.01 (.05)	0.14 (.03)
ST *n* = 71	0.42 (.06)	0.14 (.02)	0.05 (.05)	0.16 (.03)	0.08 (.04)	0.08 (.03)

Table 2.1**.** *Immediate and sustained changes in use of the misconception and accuracy for students in the control and stereotype threat (ST) learning conditions*

Change from Pretest		r^2	p	*EF* $B_{standardized}$	p	*Condition* $B_{standardized}$	p
Immediate	Misconception	.064*	.014	-.132	.126	.227**	.009
	Accuracy	.036^	.093	.182*	.038	-.074	.397
One Week Delay	Misconception	.023	.223	-.105	.230	.119	.176
	Accuracy	.039^	.076	.15^	.086	-.143	.100

Note. ^ $p < .10$, * $p < 0.05$, ** $p < 0.01$

Table 2.2. *Contributions of executive function (EF) and learning condition to immediate and sustained changes in misconception use and accuracy*

First, experiencing stereotype threat while learning predicted significantly greater immediate uptake of the misconception ($p = .009$). This suggests that those students experiencing stereotype threat while learning may have been especially focused on memorizing content from the lesson at the expense of engaging in the cognitively demanding contrast and comparison process that promotes deep learning and conceptual change. The differences reduced over time, however, most likely because the misconception use of the Control condition increased between immediate and delayed posttest, returning to closer to pretest levels. Effects were less strong in the overall sample for use of the correct solutions, with no immediate differences in accuracy but the condition differences widened over time, revealing a trend to differences in sustained learning ($p = .09$).

Student EF was positively related to immediate learning of correct content ($p = .04$) and somewhat greater ($p = .09$) sustained gains for correct content, indicating that EF may play an important role in learning from this lesson, and that high EF participants learned more overall.

EF Variations in Effects of Stereotype Threat Manipulation

While there were some impacts of condition and EF on learning for the full sample of participants, I next investigated the prediction from the testing literature, that the effects of stereotype threat are highest for high EF individuals. Condition assignment was used to predict immediate and sustained changes in misconception use and content knowledge among high and low EF students separately. All results and effect sizes from the regressions are available in Table 2.3, and mean immediate and sustained changes in misconception use and accuracy for high and low EF students in the Control and ST learning conditions are shown in Figures 2.1 and 2.2

			r^2	*Model p*	*Condition* $B_{standardized}$	*Condition* p
High EF Students	*Immediate*	*Misconception Use*	0.096	0.011	0.310	0.011
		Accuracy	0.015	0.329	-0.122	0.329
	Sustained	*Misconception Use*	0.109	0.007	0.331	0.007
		Accuracy	0.069	0.033	-0.263	0.033
Low EF Students	*Immediate*	*Misconception Use*	0.018	0.284	0.134	0.284
		Accuracy	0.000	0.977	0.004	0.977
	Sustained	*Misconception Use*	0.008	0.488	-0.087	0.488
		Accuracy	0.000	0.975	0.004	0.975

Note. $^{\wedge} p < .10$, $^{*} p < 0.05$, $^{**} p < 0.01$

Table 2.3. *Impact of stereotype threat during instruction on immediate and sustained changes in misconception use and accuracy from pretest among high vs. low executive function (EF) students*

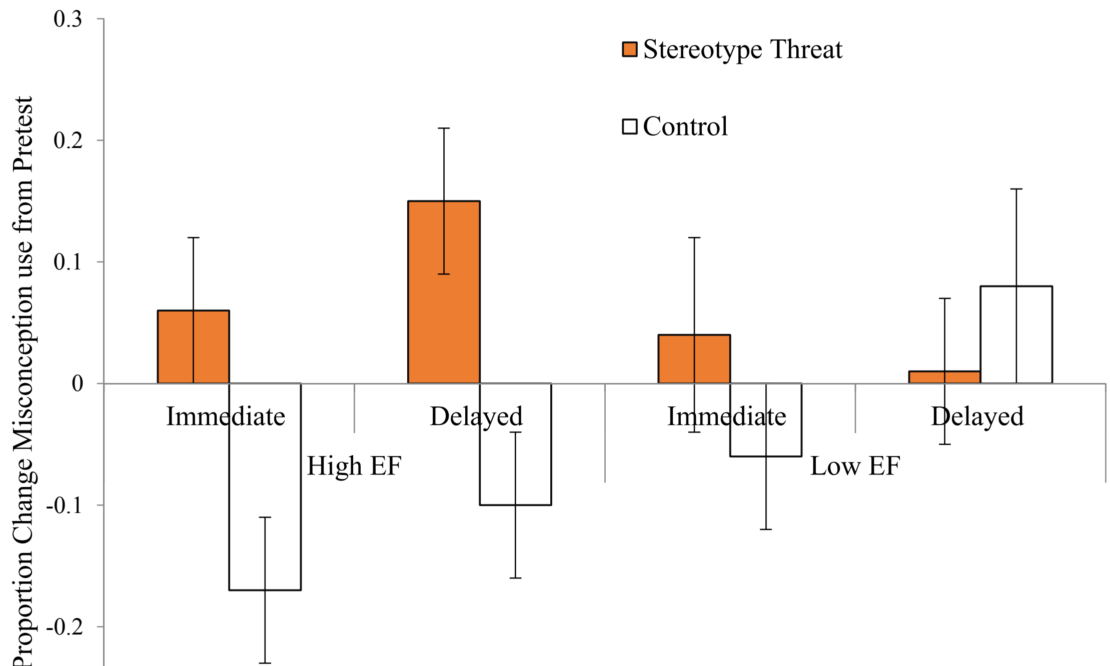


Figure 2.1 . Immediate and sustained changes in misconception use for high and low Executive Function (EF) students in the control and stereotype threat (ST) learning conditions

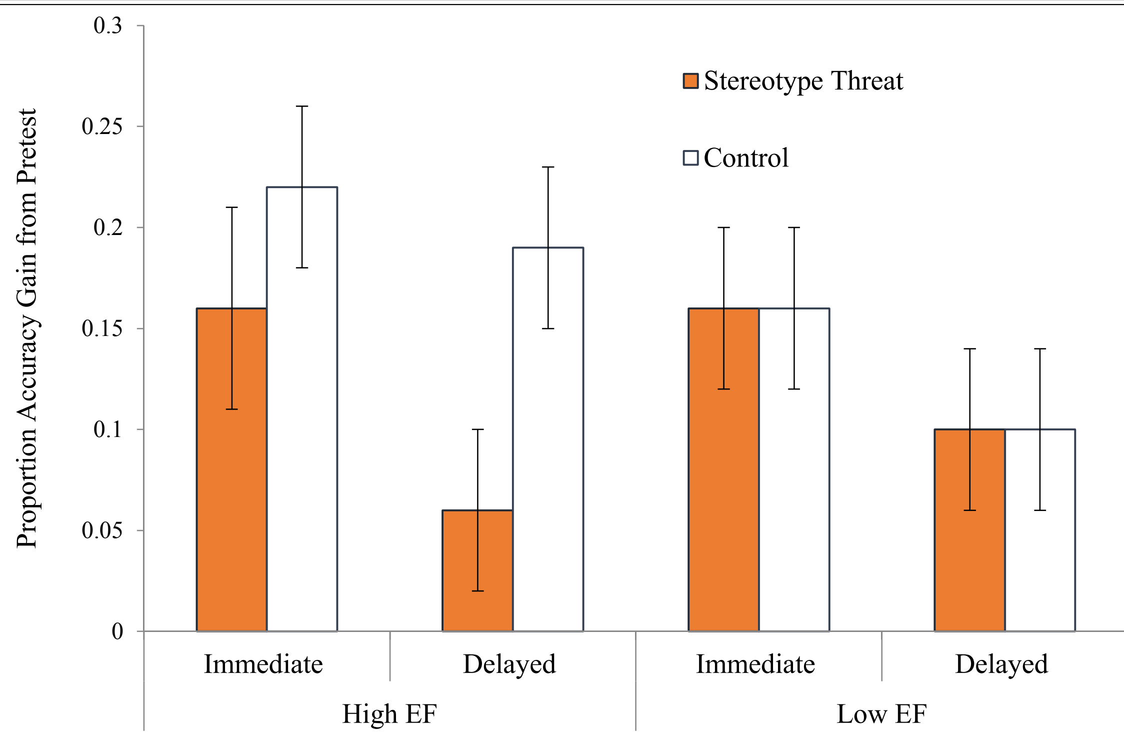


Figure 2.2. Immediate and sustained changes in accuracy for high and low executive function (EF) students in the control and stereotype threat (ST) learning conditions

As shown in Table 2.3, among high EF students, experiencing stereotype threat while learning predicted greater immediate and sustained uptake of the misconception, and smaller sustained gains in correct content, with much larger effect sizes than when considering the sample as a whole. In contrast, for low EF students, learning condition did not impact immediate or sustained gains in either correct content or misconception use.

Post hoc analyses revealed that statistically significant results all had above 0.75 observed statistical power to detect effects, with the exception of sustained gains in the high EF group (.60). Non-significant results had smaller r^2 and thereby were underpowered to detect relationships, so Type II errors of unidentified effects were possible.

Student Attitudes

In order to better understand how stereotype threat during learning impacted students' subjective experience of the lesson and likelihood of choosing to engage in a non-required math activity, I compared students' responses to the exploration intention and instant enjoyment subscales of the situational interest survey and number of optional math puzzles completed in the two conditions.

I first examined whether these subscales were related to learning. Exploration intention was related to student learning, with those students who reported greater desire to learn more about the topic showing significantly greater overall gains in accuracy, r (79) = .22, $p < .05$. In contrast, instant enjoyment did not predict overall gains in accuracy, r (79) = .22, $p = .48$.

I next assessed relations to the manipulation. A one-way ANOVA revealed that students in the ST condition reported both significantly less enjoyment than students in the Control condition (Mean Control: 4.01(0.16); ST: 3.43 (0.19), F (1, 77) = 5.22, $p = .03$; *partial* g^2 = 0.063; observed power = 0.616), and less exploration intention (Mean Control: 3.76 (0.18); ST:

3.25 (0.18), $F(1, 77) = 4.01$, $p = .05$; *partial* $g^2 = 0.048$; observed power = 0.496). These differences in students' experience of the lesson are striking because they emerged immediately after the first posttest, when performance did not yet systematically differ across conditions. Moreover, students who had experienced stereotype threat during instruction were less likely to choose to engage in optional math activities after the intervention, attempting significantly fewer math puzzles than students who had not experienced stereotype threat during instruction (Mean Control: 3.47 (0.30); ST: 2.29 (0.43), $F(1, 38) = 5.00$, $p = .03$; *partial* $g^2 = 0.116$; observed power = 0.587).

Discussion

These data provide the first evidence that stereotype threat during conceptually demanding instruction can harm mathematics attitudes and learning. Increasing the salience of race in the context of an evaluative introductory prompt led to increased misconceptions and decreased retention for the conceptual content—particularly among high EF students, who otherwise benefited most from this type of instruction. Overall, EF facilitated learning, predicting greater gains from the lesson. However, stereotype threat had a greater detrimental impact on learning among high EF students. Low EF students learned less regardless of learning condition, while greater between-condition differences were found for high EF students.

Two primary, non-exclusive, mechanisms could explain these results, and both receive some support in this study. First, it is possible that worry ideation based on a fear of confirming negative stereotypes compromised participants' EF capacity during instruction. With reduced EF available for high EF students accustomed to engaging these resources in the sorts of abstract mathematical thinking intended during this lesson, these typically higher performing, high EF students suffered the most. This is consistent with much research on pressure, anxiety, and stereotype threat, which shows the largest effects for high EF or working memory individuals (for reviews, see Beilock, 2008; Maloney et al., 2014). The ST threat may have led these students to switch from concept formation learning strategies that demand high EF resources, to less cognitively effortful memorization strategies.

A second possible contributory explanation for the results was that students who were administered the experimental prompt reported less interest and enjoyment in the mathematics. Additionally, these participants made the choice to spend less time on optional math puzzles. These findings suggest that students who experienced stereotype threat during instruction may have been less likely to think about content from the lesson between sessions or make connections with topics covered during their math classes, leading to worse consolidation of learning and resulting in the lower performance. These data were limited since they were not collected from all participants, but they suggest the potential importance of this pathway impacting both children's learning and engagement in mathematics.

This pathway received some support, as exploration intention predicted student learning, although enjoyment did not. This suggests that differences in exploration intention resulting from experiences of stereotype threat during learning may have contributed to between-condition learning differences, but that differences in enjoyment likely did not. However, even though within the boundaries of this experiment, decreased enjoyment among students who experienced stereotype threat during instruction likely did not drive between-condition learning differences, decreased enjoyment in the face of stereotype threat may have important long-term implications, possibly impacting career goals and leading to less persistence in the content domain. This mechanism is crucial to consider, because it suggests ways that perceptions of stereotype threat could have long-term implications for student engagement with a domain of knowledge, such as mathematics.

This work is a first step in understanding the phenomenon of stereotype threat during higher order thinking and learning, and it raises the possibility that subtle threats could be contributing to real and persistent achievement gaps by race, ethnicity, and gender. Importantly, whereas the effects of stereotype threat during testing are immediate but temporary and do not affect the knowledge base from which children progressively build new understandings, the effects of stereotype threat during instruction persist even when the threat is no longer present and are cumulative, as learning is built on a foundation of prior knowledge. To more fully understand similarities and differences between stereotype threat during learning versus testing opportunities, in Experiment 1b, I directly compared impacts of stereotype threat while learning with those of stereotype.

Experiment 1b

Experiment 1b replicates and extends Experiment 1a by examining impacts of stereotype threat during cognitively demanding conceptual math instruction **or** at the time of testing. Specifically, in this study, I tested how experiencing an implicit race-based academic stereotype threat during cognitively demanding conceptual math instruction or after the lesson, directly before the immediate posttest, impacted immediate and sustained learning and performance gains among African American and Latinx 5th graders.

I hypothesized that invoking stereotype threat prior to instruction could interfere with students' cognitive engagement during the lesson, leading to less sustained learning among students high in EFs, which would replicate findings from Experiment 1a. Specifically, I predicted that high EF students who were asked to report their race *before* the lesson would perform similarly to those in the control group on the immediate posttest, but exhibit no or smaller sustained learning gains at the delayed posttest, despite the threat no longer being present. This learning trajectory might indicate that students experiencing stereotype threat during the lesson were motivated to disprove the threat, and thus focused on memorizing problem solving steps and procedures, a low cognitive demand learning strategy that can maximize immediate learning gains at the expense of more enduring retention (Bjork, 1994; Christina & Bjork, 1991). Or, intrusive thoughts and worries may have loaded executive functions, limiting students' cognitive engagement with opportunities for higher order thinking and comparison embedded in the lesson to support sustained conceptual learning. Regardless of the mechanism, however, this reduction in learning would be important to understand due to the loss of an opportunity for higher order thinking and learning.

I also hypothesized that, as demonstrated in many prior studies of stereotype threat, invoking stereotype threat *after* the lesson and immediately before the test could lead students to underperform on the immediate posttest, but would not harm learning- and thus at the delayed posttest, these students would demonstrate similar gains to students in the control group.

Method

Participants

All 5th grade classes at 4 schools in the Chicago area were invited to participate. All schools invited to participate had student populations comprised nearly exclusively of students of color, who are disproportionately burdened by the challenges of contending with stereotype threat in K-12 classrooms. A total of eight fifth grade classes (193 students) participated in the study. Students who were absent on one or more study days or who reported their race to be either White or Asian were excluded from analysis, yielding a sample of 155 participants (87 girls). Students in the sample reported their race/ethnicity to be: Black or African American (55%), Latinx (29%), or Biracial (16%).

The sample was determined to be adequate to detect effect sizes where Cohen's f_2 is 0.074 or greater, which was the smallest statistically significant result obtained in Experiment 1a. In Experiment 1a, this result was obtained for immediate gains in misconception use in the full sample, $r^2 = 0.064, f^2 = 0.074$. Soper's sample size calculator (Soper, 2013), indicated the minimum sample size needed to detect effects of this size for a model with 2 predictors at a desired power of 0.80 and alpha of 0.05 to be 132. The initial and final sample were both slightly larger, allowing for uncertainty in the number of student absences.

Teacher and Experimenter Demographics

Children's regular math teachers were present in the classroom throughout the study. In all classrooms, this teacher was a white female. The primary experimenter for all classrooms was also a white female.

Design and Procedures

Students were randomly assigned within classroom to one of three conditions: Learning under stereotype Threat (LT) condition; Testing under stereotype threat (TT) condition; or Control (NT) condition. Stereotype threat was invoked using the same procedure as in Experiment 1a either: at the start of the lesson, *before* learning (LT condition); before the immediate posttest, *after* learning (TT condition) or not at all (NT condition). Procedures were otherwise identical to those in Experiment 1a.

Results

Sample

Results for 155 participants for whom I obtained complete assessment data are included in the analyses reported below (NT: n = 50, LT: n = 53, TT: n = 52). Self-reported demographics of participants at each school are shown in Table 2.4.

	Black/African American	Hispanic/Latinx	Biracial	Total
School 1 (10 girls)	0	18	4	22
School 2 (19 girls)	0	26	2	28
School 3 (24 girls)	31	1	13	45
School 4 (34 girls)	55	0	5	60
Total	86	45	24	155

Table 2.4. *Student Demographics*

Analytic Plan

As in Experiment 1a, I first report on accuracy and misconception use at pretest to ensure that random assignment was successful. I then conducted analyses testing for main effects and interactions between stereotype threat and EF on immediate and sustained learning gains. In this experiment, there were no main effects of stereotype threat for the full sample. However, the analysis detected significant interactions between learning condition and EF in predicting immediate and sustained gains in accuracy and misconception use. Thus, I next examined effects of stereotype threat among high and low EF students separately, in order to better understand this interaction and to test the hypothesis that, as was found to be the case in Experiment 1a, high EF students would be most affected by the stereotype threat manipulation. I first tested for main effects of experiencing stereotype threat at either time point. Because I found a main effect of stereotype threat among high EF students, I then examined distinct effects of stereotype threat experienced during instruction versus testing.

I operationalized students as high versus low EF based on a median split of d2 scores drawn from a corpus of data collected from 628 5^{th} grade students in the Chicago area over the past 5 years. In analyses examining impacts of stereotype threat among high EF students I include only students who scored above the median (TN-E 295) obtained in this larger data set. Relative to published norms obtained with German children aged 9 to 11, a TN-E score of 295 corresponds with a percentile rank of 72.6 (girls) and 88.5 (boys) (Brickenkamp & Zillmer, 1998). Relative to published norms obtained with German children aged 11 to 12, this score corresponds with a percentile rank of 54 (girls) and 61.8 (boys) (Brickenkamp & Zillmer, 1998). TN-E score did not differ by gender in this sample.

Performance at pretest

Neither frequency of misconception usage nor overall accuracy differed by condition at pretest (all *p*s > 0.3, Table 2.5). High and low EF students did not differ in misconception use at pretest (p =0.57), although high EF students had greater overall accuracy at pretest (p =0.004).

	Pretest Score		Immediate Gains		Sustained Gains	
	Misconception	Accuracy	Misconception	Accuracy	Misconception	Accuracy
	M (SE)	M (SE)	M (SE)	M (SE)	M (SE)	M (SE)
NT *n* = 50	0.27 (.04)	0.07 (.01)	-0.08 (.06)	0.18 (.04)	-0.09 (.04)	0.12 (.03)
ST *n* = 105	0.27 (.03)	0.09 (.01)	0.02 (.04)	0.14 (.02)	-0.01 (.03)	0.10 (.02)

Table 2.5. *Pretest Performance and Learning Gains Among Students in the Control (NT) and Stereotype Threat (ST) Conditions*

Effects of Stereotype Threat Manipulation in the Full Sample

In a hierarchical regression, I tested for main effects and interactions of Condition assignment and EF on immediate and sustained gains in accuracy and misconception use. Condition assignment (binary coded: 0 = Control; 1 =Stereotype Threat at either time point) and student EF (entered as a continuous centered variable) were first entered in the regression (Step 1) to examine main effects. Next, to test for interactions between stereotype threat and EF, a condition * EF interaction term was created and added to the model (Step 2). All results from these regressions are shown in Table 2.6. For the full sample of students, I found no main effect of stereotype threat on any of the four outcomes (Table 2.6). Effects were non-significant regardless of whether or not student gender and school were included in the model as control variables. I repeated these same analyses, testing for effects of stereotype threat experienced during versus after instruction separately and found that, for the full sample of students, neither stereotype threat during versus after instruction predicted immediate or sustained gains in misconception use or accuracy. These results diverge slightly from findings in Experiment 1a, which showed trends towards worse learning under stereotype threat in the full sample of students- although these trends were driven entirely by high EF students.

	Learning Gains							
	Immediate				Sustained			
	Misconception		Accuracy		Misconception		Accuracy	
Predictor	**Δ** R^2	*B*	**Δ** R^2	*B*	**Δ** R^2	*B*	**Δ** R^2	*B*
Step 1	0.005		0.02		0.02		0.01	
ST		0.07		-0.08		0.11		-0.05
EF		-0.01		0.12		0.05		0.08
Step 2	0.04*		0.04*		0.04*		0.04**	
ST x EF		0.31*		-0.31*		0.30*		-0.34*
Total R^2	0.04*		0.06*		0.05*		0.05**	

Note. All effect sizes are standardized

Note. ^ $p < 0.10$. * $p < 0.05$. ** $p < 0.01$

Table 2.6. *Multiple Regression Analyses Predicting Immediate and Sustained Learning Gains From Main Effects and Interactions of Stereotype Threat (Before Learning Or Testing) and EF*

However, while there were no main effects of stereotype threat, the analysis revealed significant interactions between learning condition and student EF for each of the outcomes examined (Table 2.6). To better understand these interactions, I next examined effects of stereotype threat on learning among high and low EF students separately.

EF Variations in Main Effects of Stereotype Threat Manipulation

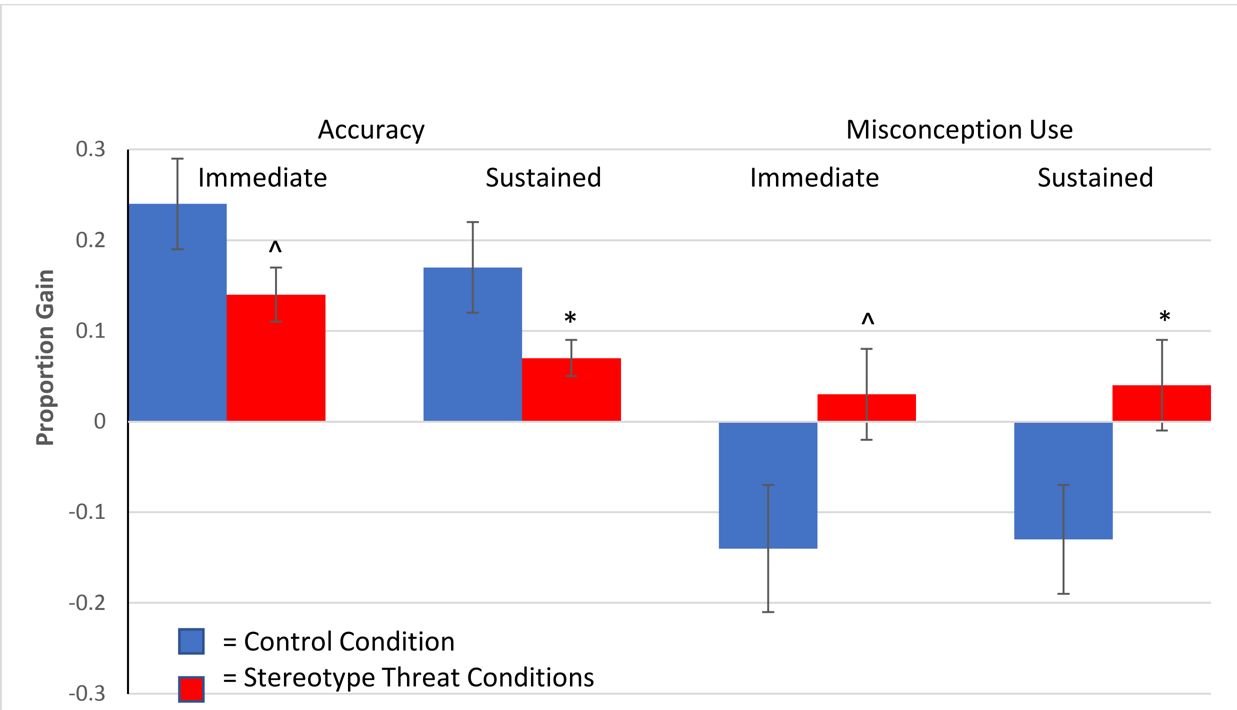
I next examined effects of stereotype threat among high and low EF students separately in order to test the prediction and replicability of the finding in Experiment 1a that detrimental effects of stereotype threat are concentrated among high EF students. I first tested for main effects of experiencing stereotype threat at either time point. Condition assignment (binary coded: 0 = Control; 1 = Stereotype Threat at either time point) was used to predict immediate and sustained changes in misconception use and accuracy among high and low EF students separately. High EF students who experienced stereotype threat at either time point had smaller sustained gains in correct content ($B_{standardized}$ = 0-.236, p = 0.026), and greater sustained misconception uptake ($B_{standardized}$ = 0.220, p = 0.038), with additional trends towards worse immediate learning (Table 2.7).

			R^2	ST B	Exact p value
High EF Students	Immediate	Misconception Use	0.040^	0.200^	0.060
		Accuracy	0.042^	-0.204^	0.055
	Sustained	Misconception Use	0.048*	0.220*	0.038
		Accuracy	0.056*	-0.236*	0.026
Low EF Students	Immediate	Misconception Use	0.013	-0.115	0.357
		Accuracy	0.010	0.100	0.426
	Sustained	Misconception Use	0.005	-0.068	0.586
		Accuracy	0.029	0.171	0.169

Note. ^ $p < 0.10$. * $p < 0.05$, ** $p < 0.01$.

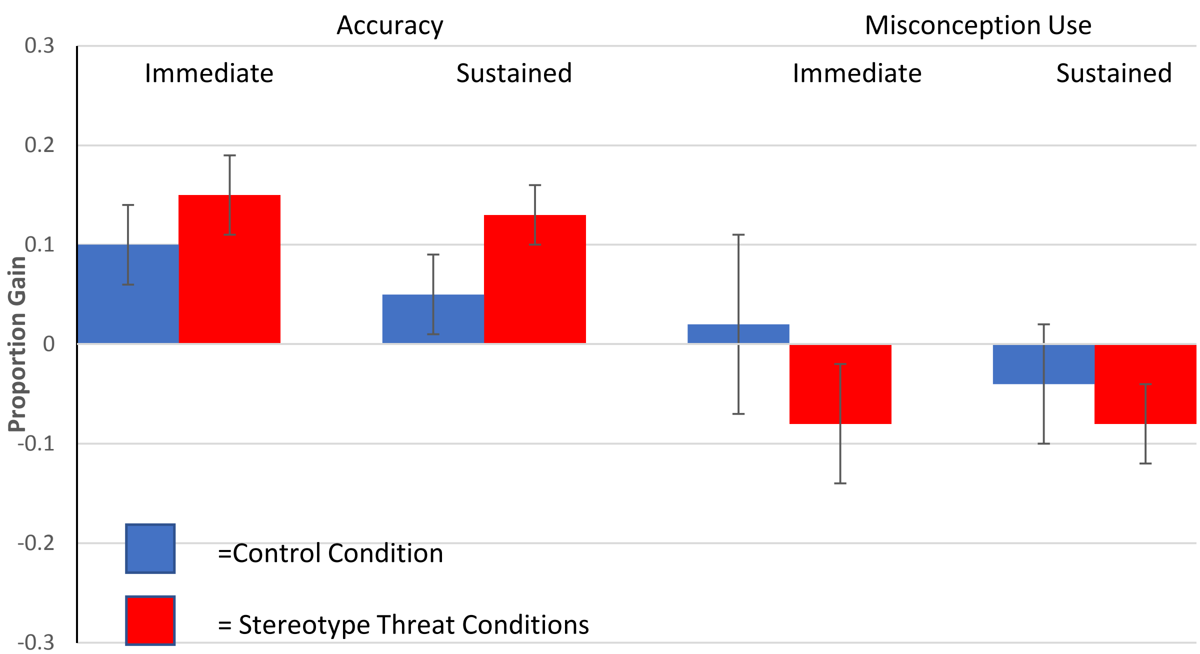
Table 2.7. Regression Analysis Predicting Immediate and Sustained Learning Gains From Stereotype Threat (Before Learning Or Testing) among High vs. Low EF Students

Main effects of stereotype threat among high EF students on sustained gains for both accuracy and misconception usage are robust and remain significant whether or not the following control variables are included in the model: student gender, school, pretest performance. In contrast, among low EF students, stereotype threat did not predict learning gains on any of the outcomes (Table 2.7). Effects remain non-significant when controls are included in the model. Mean immediate and sustained changes in misconception use and accuracy for high and low EF students in the Control and stereotype threat learning conditions are shown in Figures 2.1. and 2.2.



^ $p < 0.10$. * $p < 0.05$

Figure 2.3. Immediate and Sustained Learning Gains Among High EF Students in the Control and Stereotype Threat Learning Conditions



^ $p < 0.10$. * $p < 0.05$

Figure 2.4. Immediate and Sustained Learning Gains Among Low EF Students in the Control and Stereotype Threat Learning Conditions

Because I found a main effect of stereotype threat among high EF students, I next examined effects of stereotype threat experienced before learning versus testing separately for this group.

Stereotype Threat during Learning

I first tested for distinct effects of stereotype threat experienced while learning among high EF students. I repeated the same analysis as above in order to compare learning among students assigned to the LT and Control conditions. All results and effect sizes from the regression are shown in Table 2.8. I found that, although immediate learning gains among students in the LT condition did not differ significantly from those of students in the Control condition, students assigned to the LT condition had significantly smaller sustained gains in correct content ($B_{standardized}$ = -0.250, p = 0.043) and somewhat greater sustained misconception usage ($B_{standardized}$ = -0.250, p= 0.052). When school and gender were included in the model, impacts on sustained misconception usage attained significance (R square change = 0.06, $B_{standardized}$ = -0.250, p = 0.04.)

		R^2	LT B	Exact p value
Immediate	Misconception Use	0.024	0.154	0.236
	Accuracy	0.035	-0.187	0.148
Sustained	Misconception Use	$0.063^{\wedge}$	$0.250^{\wedge}$	0.052
	Accuracy	0.068^{*}	-0.260^{*}	0.043

Note. All effect sizes are standardized

^ $p < 0.10$. * $p < 0.05$.

Table 2.8. Regression Analysis Predicting Immediate and Sustained Learning Gains From Stereotype Threat Before **Learning** (LT) Among High EF Students

This pattern of results in which the detrimental impact of experiencing stereotype threat during learning is most evident at a delay replicates findings in Experiment 1a and suggests that intrusive thoughts from stereotype threat may have interfered with the deep engagement and conceptual learning needed for sustained retention of new content. Relative to students in either of the other two learning conditions, students assigned to the LT condition experienced striking declines in accuracy and increases in misconception usage between the immediate and delayed posttests (Figures 2.5 and 2.6).

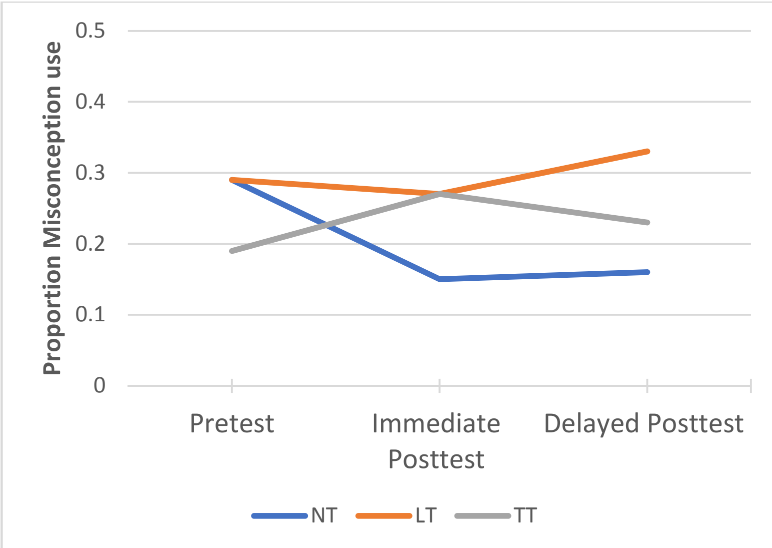


Figure 2.5. Misconception Use Among High EF Students in the Control (NT), Learning Threat (LT) and Testing Threat (TT) Experimental Conditions

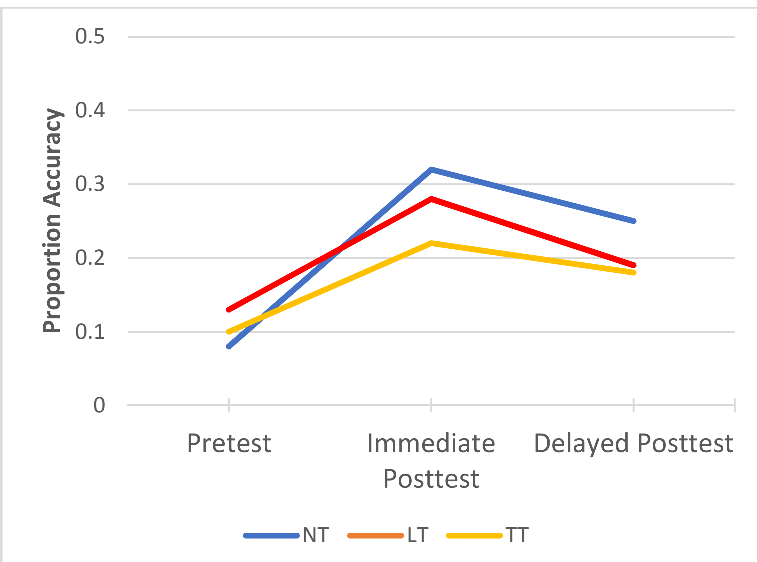


Figure 2.6. Accuracy Among High EF Students in the Control (NT), Learning Threat (LT) and Testing Threat (TT) Experimental Conditions

Stereotype Threat during Testing

I next examined effects of stereotype threat during testing, repeating the identical analyses. In contrast to stereotype threat experienced during learning, effects of stereotype threat while testing were evident immediately and reduced over time. As shown in Table 2.9, being assigned to the TT condition predicted larger immediate uptake of the misconception and trends towards smaller immediate gains in correct content. Although effects reduced over time, learning among students assigned to the TT condition did not rebound fully to that of students in the control condition (Figures 2.3 and 2.4).

		R^2	TT *B*	Exact *p* value
Immediate	Misconception Use	0.085*	0.291*	0.028
	Accuracy	0.051^	-0.226^	0.090
Sustained	Misconception Use	0.050^	0.224^	0.093
	Accuracy	0.040	-0.201	0.134

Note. All effect sizes are standardized

^ $p < 0.10$. * $p < 0.05$.

Table 2.9 Regression Analysis Predicting Immediate and Sustained Learning Gains From Stereotype Threat Before **Testing** (TT) among High EF Students

Discussion

Worries and concerns about being judged through the lens of negative stereotypes can characterize everyday learning experiences for many students from academically stigmatized groups (Nasir, 2017, D' hondt, Eccles, Houtte, & Stevens, 2016). Stereotypes can be especially salient in mathematics contexts, a subject area in which success is often perceived as being dependent on innate talent or brilliance (Chestnut, Lei; Leslie& Cimpian, 2018). These worries and concerns can have real consequences for students' learning. In the series of studies reported in this chapter, an implicit prompt to highlight identity and thereby potentially invoke stereotype threat, either before or after conceptual mathematics instruction, harmed achievement among students high in EFs. In contrast, low EF students' achievement was not impacted by stereotype threat at either time point. This pattern of results indicates that stereotype threat may be selectively harming learning and achievement among those very individuals with the greatest potential to excel within and contribute to the stereotyped domain, potentially contributing to what has been referred to as the "excellence gap" (see Plucker, et al., 2010).

Additionally, in line with the concerning possibility raised in the discussion for Experiment 1a that effects of stereotype threat while learning may be more long-lasting than those of stereotype threat during testing, findings from Experiment 1b indicate that the role of stereotype threat in shaping initial learning differs from its role in performance contexts. While stereotype threat at either time point was harmful, when stereotype threat was experienced prior to instruction, effects were especially enduring. High EF Students who experienced stereotype threat during instruction showed the largest learning decrements at the delayed posttest, despite the threat no longer being present. In contrast, when stereotype was invoked after learning and immediately before the test, achievement decrements among high EF students were evident immediately and reduced over time. These findings indicate that stereotype threat in classroom contexts, whether experienced during learning or testing opportunities can meaningfully impact achievement among students high in EFs. Additionally, these findings suggest that impacts of stereotype threat while learning differ from those observed in testing contexts, underscoring the need for additional research on stereotype threat during opportunities for initial learning.

In Experiment 1b, although high EF students in the Learning Threat condition exhibited similar learning gains as those in the control condition immediately following the lesson, they retained less learning from the lesson. This learning trajectory among high EF students experiencing stereotype threat during instruction replicates findings from Experiment 1a. These results may be because the cognitive load from experiencing stereotype threat while learning led students to engage with the lesson on a more superficial level- focusing on memorizing content and procedures, at the expense of deep engagement with novel concepts. For example, a student who focused his attention during the lesson on memorizing the procedures for solving ratio problems may have performed similarly to (or even better than) one who focused her energy on understanding the conceptual underpinnings of the lesson, but would be more likely to forget the information over time. An additional possible explanation for this achievement trajectory should be considered as well, which is that students in the learning threat condition could have actually truly learned similar amounts during the lesson itself, but been motivated to forget content from the lesson, which they associated with an unpleasant experience, a phenomenon that may occur in response to stressful classroom contexts (see Ramirez, 2017). Additional research is needed to disentangle these possibilities. Research that closely examines pathways and mechanisms through which experiences of stereotype threat during instruction can lead to less sustained learning among students with the greatest potential could help to inform efforts to ameliorate these detrimental effects.

To date, much policy and research on racialized achievement gaps has focused on supporting struggling students from historically under-performing groups to reach proficiency. However, the education research field has increasingly recognized the need to also attend to challenges faced by academically successful students of color, in order to reduce racialized patterns of representation at the upper ends of the achievement spectrum. Indeed, in a large-scale, longitudinal study Reardon (2008) found that the black-white achievement gap grew most quickly among students who were performing above the median in reading and math at kindergarten entry. Additionally, many of the most pronounced racial inequities are evident in enrollment patterns in gifted and talented programs (Grissom, & Redding, 2016). One factor that has been suggested as contributing to achievement gaps at the upper end of the achievement spectrum is that high achieving black students may have access to fewer challenging learning opportunities (Reardon, 2008). Results from this study suggest that experiences of stereotype threat may also reduce cognitive engagement during conceptually challenging instruction, even when these opportunities are available, and highlight the importance of working to prevent experiences of stereotype threat in learning as well as testing contexts. Understanding and addressing factors shaping achievement gaps among high-achieving students will be a powerful leverage point towards increasing the diversity of STEM fields.

CHAPTER THREE: EVALUATIVE PERFORMANCE PRESSURE DURING LEARNING AND TESTING

Introduction

In Experiments 1a and b, identity threatening pressure during cognitively demanding math instruction acted as a threat, reducing learning and interest. While the size of the detrimental impact depended on students' baseline EFs, there was no evidence that stereotype threat acted as a motivating incentive and improved outcomes for any subset of students.

In this Chapter, I describe impacts of experiencing evaluative performance pressure, heightened pressure not explicitly linked to identity, within the context of the same cognitively demanding math lesson. Specifically, I tested effects of an evaluative performance pressure manipulation in which students were told that their performance would determine whether or not their entire class would receive a desired incentive. As with stereotype threat, evaluative performance pressure can result in intrusive thoughts that interfere with cognition (Beilock & Carr, 2005; Schmader & Johns, 2003; Schmader, Johns & Forbers, 2008). However, I hypothesized that because identity is not explicitly implicated, evaluative performance pressure might have greater potential to also act as a motivator. Among some students, it could improve learning through increased focus, while for others it could harm learning through intrusive thoughts and cognitive load.

Among which students does evaluative performance pressure help vs. harm learning? In order to identify student level factors that might shape whether pressure helps vs. harms learning, I tested whether impacts of the evaluative performance pressure manipulation were moderated by either student's baseline EFs or gender. As elaborated upon in the introduction, gender and EF have been found to moderate effects of pressure experienced while testing. The largest performance *decrements* under pressure are often seen among high EF individuals, which may be because these individuals are more likely to rely on cognitively demanding strategies to solve problems when not experiencing pressure (Maloney, Sattizahn, & Beilock, 2014; Maloney, Schaeffer& Beilock, 2013). On the other hand, in cases where pressure acts more as a motivating incentive than as a distracting threat, the largest performance *gains* under pressure are often seen among males (Attali, Neeman &Schlosser, 2010; Levitt, List, Neckerman & Sadoff, 2016). This has been hypothesized as possibly stemming from gender differences in time sensitivity to rewards or to males putting forth less effort in low pressure settings (e.g. Attali, Neeman &Schlosser, 2010; Levitt, List, Neckerman & Sadoff, 2016).

Pressure and Optimal Learning and Performance

Although pressure and stress are often viewed through a negative lens as experiences to be avoided, human performance actually peaks under conditions of moderate stress and arousal (see Yerkes & Dodson, 1908; Sapolsky, 2015). Indeed, across many dimensions of human functioning, outcomes improve during the transition from a complete absence of stress to mild stress; however, as stress becomes more severe, outcomes plateau and eventually decline (see Sapolsky, 2015). Likewise, we would expect children to learn best under conditions of moderate stress and arousal- when they are motivated to do well, but not overly stressed out.

In the case of mathematics, when pressure is too low, learning and performance may be harmed due to inadequate task engagement and effort (Sullivan, Tobias & McDonough, 2006; Martin et al, 2012). Alternatively, too much pressure and anxiety can also interfere with learning (and performance. Indeed, middle school students encounter challenges to mathematics success due to both too much (i.e. worry and anxiety) and too little (i.e. disengagement) pressure.

On the one hand, too much pressure can pose a challenge to math success if intrusive thoughts and worries tax cognitive resources, limiting students' cognitive engagement during math learning and testing opportunities (e.g. Lyons, Simms, Begolli & Richland, 2018; Maloney, Sattizahn, & Beilock, 2014). Students may experience harmful pressure and anxiety in math contexts due to situational factors (for example, pressure associated with high stakes testing), as well as stable characteristics or traits (for example, math anxiety). Trait math anxiety is associated with low math knowledge and grades, as well as less adaptive motivational frameworks (Ashkraft & Krause, 2007; Gunderson, Park, Maloney, Beilock & Levine, 2018). Additionally, situational factors, such as experimental manipulations of heightened pressure have been shown to result in decrements in math performance (for review, see Beilock, 2008) and learning (e.g. Lyons, Simms, Begolli & Richland, 2017). As math achievement is often high stakes, for example being used in course placement, tracking and grade promotion decisions, many students feel a great deal of pressure during everyday math learning opportunities, which could interfere with math learning.

On the other hand, if students do not see mathematics as having personal relevance within their current or future lives, they may feel little motivation to learn or perform well and may disengage during math learning and testing opportunities. Thus, while some students may feel so much pressure while doing math that anxious ideation interferes with their learning and performance, a significant portion of students may experience poor learning and performance outcomes due to limited task engagement and effort. Indeed, it has been observed, both in the US and internationally, that a significant number of students are disengaging or “switching off” in math (Sullivan, Tobias & McDonough, 2006; Martin Anderson, Bobis, Way& Vellar, 2012), with the middle school years being identified as a time period during which boys in particular may be especially susceptible to disengagement form math (Gottfried, Marcoulides, Gottfried, Oliver & Guerin, (2007;Martin & Marsh, 2005). For these students, too little stress or arousal may contribute to disengagement, thus posing a challenge to their mathematics success.

As too much (e.g. anxiety) and too little (e.g. disengagement) pressure can both pose obstacles to math success during middle school, raising the stakes of learning and performance has the potential to both help and hinder math achievement. On the one hand, heightened pressure could act as a distracting threat, interfering with cognitive engagement. On the other hand, increased pressure could act as a motivating incentive- providing a nudge towards greater effort and engagement.

Because optimal learning can occur when students are motivated, but not overly stressed out, ensuring that mathematics instructional time is spent under optimal levels of pressure could be a powerful leverage point for improving mathematics achievement. However, exactly how to do this is unclear due to the potential for increasing the pressure students experience while doing math to have positive as well as negative impacts depending on students' initial stress and engagement during math.

Despite extensive research demonstrating a non-linear relationship between stress and human functioning across many outcomes (for review, see Sapolsky, 2015), research in education contexts has mostly focused either on the role of pressure as a motivating incentive *or* on pressure as a distracting threat, with very little research engaging at once with the potential for pressure to act as both support and challenge to math success. This research, which is reviewed briefly below, has found both positive and negative effects of raising the stakes of performance for student outcomes.

Pressure as motivating incentive. Research, predominantly in the field of behavioral economics, that has focused on the incentivizing potential of increased pressure has demonstrated that rewarding students' effort and performance can be effective in boosting achievement outcomes. Many studies have focused on the impact of financial incentives, documenting that paying students can result in better performance (e.g. Leuven, Oosterbeek & Klaauw, 2010; List, Livingston, & Neckermann, 2018.) However, it appears that non-financial incentives may be equally effective, at least among younger children. For example, in a large randomized field experiment conducted with students in the Chicago area, Levitt, List, Neckerman & Sadoff (2016) found that both financial and nonfinancial incentives improved test performance among elementary school students. The effects were somewhat larger for boys than for girls, which the authors suggest may be due to boys' greater sensitivity to short-term incentives (Levitt at al., 2016).

The prospect for increased pressure to lead to better performance is also supported by findings that GRE performance is higher under pressure (i.e. the real GRE) than in a low stakes situation (i.e. a voluntary experimental section of the GRE that participants could select to take following the actual exam) (Attali, Neeman & Schlosser, 2011). The performance advantage in the high stakes situation was larger for males than for females, which the authors suggest may be due to males putting in less effort in the low stakes situation (Attali et al., 2011).

Pressure as distracting threat. While the possibility of being rewarded or sanctioned for performance can be motivating, it can also be distracting. In contrast to the behavioral economics literature, research in the fields of psychology and education has mostly explored the downsides of pressure. Studies conducted in laboratory as well as classroom contexts demonstrate that experimentally invoking increased pressure can in many cases harm performance, likely due to intrusive thoughts and worries that tax EF resources (Schmader & Beilock, 2012), interfering with task engagement (DeCaro et. al., 2011). In addition to general worries about performance, it is also possible that increased pressure and focus on evaluation may lead females and minorities to worry that their performance will be judged through the lens of negative stereotypes (Steele & Aronson, 1995).

Gender and Mathematics

Gender gaps in mathematics achievement in high and low stakes settings.

Mathematics achievement and eventual career trajectories continue to be patterned by gender in complex and oftentimes paradoxical ways, ultimately resulting in math-intensive fields remaining heavily male dominated (National Science Foundation, 2017). These patterns are not simple achievement gaps, however, since measures of mathematics achievement requiring sustained effort in low-stakes settings often favor girls, while the gender gap reverses on high-stakes measures of math achievement such as standardized tests and mathematics competitions. For example, girls tend to earn higher grades than their male counterparts across subjects including math throughout middle and high school (e.g. Easton, Johnson & Sartain, 2017), invest more time in homework (e.g. Gershenson & Holt, 2015) and are more likely to persist to high school and college graduation (Stetser & Stillwell, 2014). Test scores, however, tell a different story: although gender gaps in average math achievement have reduced such that in most cases males and females perform similarly on standardized assessments of math achievement, where gender gaps in average achievement do exist, they tend to favor males (Reardon et al., 2018). In those cases, gender gaps vary across districts, with boys particularly continuing to exhibit higher average mathematics performance in wealthier school districts (Reardon et al., 2018). Additionally, even though gender gaps at average levels of performance have shrunk over time, gender gaps at the highest levels of mathematics achievement remain large. For example, consistently less than 10 % of 9^{th} to 12^{th} grade students scoring in the Top 50 level of achievement on the American Mathematics Competition are female (Ellison& Swanson, 2018).

Gender as risk and protective factor for mathematics success. In math contexts, females may worry that their learning and performance will be judged based on negative stereotypes (Shapiro, 2012; Spencer, Steele & Quinn, 1999). In addition to contending with negative stereotypes about their mathematics ability, girls and women are also more likely to exhibit high math anxiety (Hembree, 1990; Devine et al., 2012). Although math anxiety is a powerful predictor of low math achievement across diverse contexts(see Chang & Beilock, 2016; Foley et al., 2017), even when studies find gender differences in math anxiety they do not always find gender differences in math achievement. For example, in an adolescent sample (average age 12-15), Devine et al. (2012) found that girls had higher math anxiety and that high levels of math anxiety predicted lower math achievement. However, despite girls' higher levels of math anxiety, their performance was equal to that of boys. The authors suggest these findings indicate that girls may have the potential to outperform boys in mathematics, if not for their math anxiety (Devine et al., 2012).

As mathematical reasoning develops "from a set of biologically based cognitive capacities that males and females share" (see Spelke, 2005), it becomes important to consider what protective factors might be contributing to girls' relative success in math, despite their higher math anxiety and the persistence of negative stereotypes about girls and math. Or, alternatively, what risk factors may pose challenges to boys' math achievement, despite their relatively low math anxiety and the persistence of positive stereotypes about boys and math. One important protective factor for girls- or alternatively risk factor for boys- may be academic engagement. In elementary and middle schools, boys often exhibit lower levels of behavioral and emotional academic engagement than do girls, which could be a contributor to boys' relative declines in academic achievement during the middle school years (Kindermann, 2007; Marks, 2000).

The gendered patterns of mathematics achievement described above, specifically the discrepancy in the directionality of gender gaps between high and low stakes measures of math achievement, suggest that the role of pressure should be more meaningfully considered in this context. One reason we might expect to see gender differences in the role of pressure in shaping mathematics achievement is that females tend to have higher mathematics anxiety (Hembree, 1990; Devine et al., 2012). If females are starting out more anxious, introducing additional pressure would be more likely to push them outside the zone of optimal learning and performance- and into the zone of too much stress, where cognition suffers. In contrast, if males begin with relatively low anxiety, an added source of pressure would be more likely to push them into the zone of optimal learning and performance. Another reason that pressure could function differently for males and females is that concerns about being judged stereotypically could be intensified in the context of increased pressure and focus on evaluation.

Methods

Participants

Study participants were 5th grade students drawn from 5 schools in the Chicago area. Participating schools included 1 traditional public school (School 1), 2 Catholic schools (Schools 2 and 3), 1 charter school (School 4), and 1 private school (School 5). Permission was obtained to conduct research in all 5th grade classrooms at participating schools (8 classes total). A total of 205 students participated. Twenty-seven students who were absent on one or more study days were excluded due to missing data, leaving 178 students. A breakdown of student demographic information by school is shown in Table 1.

Teacher and Experimenter Demographics

Children's regular math teachers were present in the classroom throughout the study. In l 7 classrooms, the teacher was a white female. In one classroom at school 5, the teacher was an African American male. The primary experimenter in all classrooms was a white female.

		Hispanic	African American	White	Biracial	Total
School 1: Traditional Public	Boy	2	20	0	1	23
	Girl	0	15	1	5	21
School 2: Catholic	Boy	0	0	7	1	8
	Girl	0	0	5	0	5
School 3: Catholic	Boy	1	6	1	2	10
	Girl	4	2	0	2	8
School 4: Charter	Boy	21	0	2	6	29
	Girl	17	0	3	7	27
School 5: Private	Boy	0	5	9	4	18
	Girl	0	5	16	8	29
Total		45	53	44	36	178

Table 3.1. Student demographics

Design and Procedures

Study procedures and design were identical to those of Experiment 1b, with the exception of the particular pressure manipulation. As in Experiment 1b, students were randomly assigned within each classroom to experience heightened pressure during learning (LP condition), during testing (TP condition), or not at all (NP condition).

In this study, I examined impacts of a non-identity threatening source of pressure: evaluative performance pressure. The pressure manipulation used in this study was modeled on the pressure manipulation used in Beilock et al. (2004), which effectively induced feelings of pressure and social evaluative threat by informing participants that their performance would determine not only whether or not they would receive a reward, but also whether or not a partner would receive the reward. To most closely mimic the types of pressure conditions experienced in schools, where rewards and sanctions are often delivered on a class or even school-wide level, in this study, students were told that whether or not their entire class would receive a reward was dependent on their performance.

Specifically, students in the pressure conditions were told, either before learning (LP condition) or before testing (TP condition), that they would be taking a test, and if they scored at least 80%, their class would be given a pizza party, but if they failed to earn 80% or higher, their class would lose the pizza party. Specifically, students viewed a screen in which the following prompt was made visible and read aloud either before (LP condition) or after (TP condition) the lesson:

Thank you for taking part in this study. We're glad that you are doing this activity because it will help us learn about some of the best ways to teach kids math. To show our appreciation for your helping us out with our study, we are hoping to give your class a pizza party.

You will be taking a test after the lesson. To make sure that people in your class pay attention to the lesson, we have selected some students to decide whether or not your class will get the pizza party at the end. You are one of those students! If ***ALL*** *the selected students score above 80% on the test, your class will have the pizza party. If* ***ANY*** *of these students get less than 80% on the test, your class will not have the pizza party. So, your class is counting on you to score your best. If you do not earn at least 80% on the test, your class will lose the pizza party.*

In contrast, students in the no pressure condition were told the aim of the study was to better understand how students learn math and that after the lesson they would be asked to solve some problems. Specifically, students viewed a screen in which the following prompt was made visible and read aloud:

Thank you for taking part in this study. We're glad that you are doing this activity because it will help us learn about some of the best ways to teach kids math. We appreciate your helping us out with this study! Once we've finished the study, we will share what we find out about how kids learn math with your class.

After the lesson, we will ask you to solve some problems and answer some questions about what you've learned. Please pay attention to the lesson. We hope that the lesson will help you to understand important math concepts more deeply! Understanding math concepts deeply will be important for middle school and high school math.

Results

Analytical Plan

In the analyses that follow, I first describe pretest performance for students assigned to the 3 conditions, in order to confirm success of random assignment. In the description of pretest performance, I also report on student level factors that predicted pretest performance.I next describe gender differences in student learning gains across the No Pressure, Learning Pressure and Testing Pressure experimental conditions, and test whether impacts of pressure during learning or testing differed for boys and girls. Finally, I describe gender differences in student engagement in the high vs. low pressure experimental conditions, and test whether the role of pressure during learning or testing in shaping engagement differed for boys and girls.

Pretest Performance

Pretest performance did not differ between students assigned to the 3 conditions or between boys and girls (all *p*s >0.24). Mean pretest scores among boys and girls assigned to each of the 3 learning conditions are shown in Table 3.2. Pretest performance was not predicted by student EF or race, but did differ between schools. A dummy variable for student school, along with pretest performance, were included as controls in all analyses of learning outcomes.

	Pretest Performance	Immediate Gains	Sustained Gains
Boys	Mean (SE)	Mean (SE)	Mean (SE)
NP (n=26)	0.25 (0.06)	0.13 (0.04)	0.12 (0.04)
LP (n=33)	0.21 (0.05)	0.16 (0.04)	0.18 (0.05)
TP (n=30)	0.20 (0.05)	0.23 (0.06)	0.22 (0.06)
All (n=89)	0.21 (0.03)	0.18 (0.03)	0.18 (0.03)
Girls	Mean (SE)	Mean (SE)	Mean (SE)
NP (n=28)	0.24 (0.05)	0.30 (0.07)	0.24 (0.06)
LP (n=30)	0.21 (0.04)	0.22 (0.04)	0.18 (0.04)
TP (n=31)	0.28 (0.05)	0.24 (0.05)	0.20 (0.05)
All (n=89)	0.25 (0.03)	0.25 (0.03)	0.20 (0.03)

Table 3.2. Mean pretest performance and learning gains among boys and girls assigned to the No Pressure (NP), Learning Pressure (LP) and Testing Pressure (TP) Conditions.

Gender Gaps in learning in the high vs. low pressure experimental conditions

I first conducted a series of regressions to examine predictors of student learning in the absence of pressure, among the 54 students (28 girls) assigned to the No Pressure (NP) condition. Pretest performance and a dummy variable for school were first entered into the regression (Model 1), as control variables. Dummy variables for each of the 5 schools were created, coded as 1 if the child was from that school; 0 otherwise. The variable for the school with the greatest number of students, school 4, was not entered into the regression. Next, student gender, race/ethnicity and EF were added to the model (Model 2). Student gender was entered as a dummy variable (1= female, 0 = male). Dummy variables for student race/ethnicity were included in the model as follows: (African American; 1 if African American, 0 otherwise; Latinx: 1 if Latinx, 0 otherwise; White: 1 if White, 0 otherwise; Biracial: 1 if Biracial, 0 otherwise). As with school, the variable for the student/race ethnicity with the greatest number of participants, African American, was not entered into the regression. Student EF was entered as a continuous variable with values ranging form 132 to 444. Results of the full regression are shown in Table 3.3. Neither student EF nor race/ethnicity predicted learning outcomes, although it is possible that the model may have been underpowered to detect these relations.

The only student characteristic that predicted either immediate or sustained learning gains among students in the no pressure condition, after controlling for school and pretest performance, was gender. Girls exhibited significantly larger immediate learning gains ($\beta_{standardized} = 0.29$, $p = 0.03$), and trends towards greater sustained learning gains ($\beta_{standardized} = 0.21$, $p = 0.09$). I also repeated the exact same analysis to examine predictors of student learning in the pressure conditions; gender did not predict learning gains among students in either the Learning Pressure or Testing Pressure conditions (all p s > 0.45). Mirroring broader-scale patterns of achievement, girls had larger learning gains in the low pressure context (Figure 3.1), while the gender gap disappeared, and showed possible trends towards reversing, in the high pressure conditions.

	Immediate Gains		Sustained Gains	
Predictor	**Δ** R^2	*B*	**Δ** R^2	*B*
Step 1	0.25^{*}		0.17^{*}	
Pretest		-0.45^{**}		-0.38^{**}
School 1		-0.20		-0.18
School 2		-0.20		-0.20
School 3		$-0.25^{\wedge}$		-0.11
School 5		0.05		0.04
Step 2	0.16^{*}		0.18^{*}	
Gender		0.29^{*}		$0.21^{\wedge}$
EF		0.10		0.21
White		0.42		0.41
Latinx		-0.08		0.03
Biracial		-0.06		-0.02
Total R^2	0.41		0.40	

Note. All effect sizes are standardized

^ p <0.10 * p <0.05. ** p <0.01.

Table 3.3. Predictors of Student Learning in Low Pressure Context

Figure 3.1a . Learning Gains: No Pressure Condition

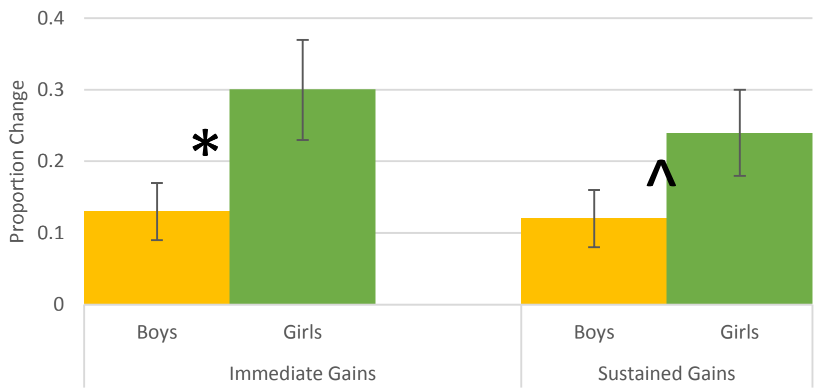


Figure 3.1.b Learning Gains: Learning Pressure (LP) Condition

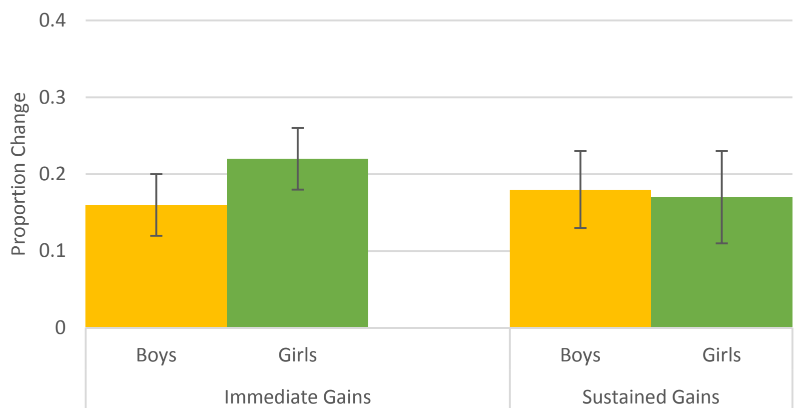
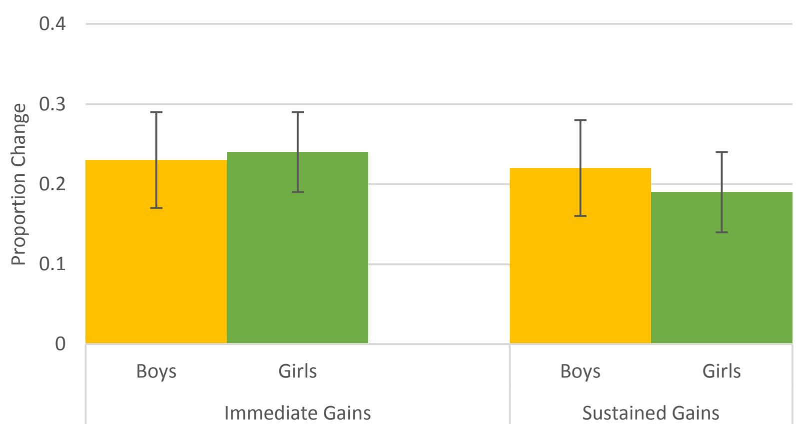


Figure 3.1. c Learning Gains: Pressure (TP) Condition



Gender differences in impacts of pressure during learning on learning outcomes

To examine whether gender differences in the role of pressure in shaping learning might help explain the differences in gender gaps in math across high vs. low pressure learning contexts, I next examined main effects and interactions of the Learning Pressure study manipulation and student gender.

Pretest score, along with a dummy variable for school were first entered into the regression (Step 1), as control variables. Main effects (Student gender, LP study manipulation) were added at step 2. Finally, to test the possibility that the role of heightened pressure in shaping learning differed for boys and girl in this study, a pressure * gender interaction term was added to the analysis (Model 3). Results of the full regression are shown in Table 3.4. The analysis indicated that gender interacted with the LP study manipulation to predict sustained learning gains ($\beta_{standardized} = -0.32$, $p = 0.05$) and may have also interacted to predict immediate learning gains ($\beta_{standardized} = -0.27$, $p = 0.07$).

	Immediate Gains		Sustained Gains	
Predictor	**Δ** R^2	*B*	**Δ** R^2	*B*
Step 1	0.21^{***}		0.14^{*}	
Pretest		-0.42^{***}		-0.35^{***}
School 1		-0.21^{*}		-0.16
School 2		$-0.17^{\wedge}$		-0.11
School 3		-0.23^{*}		-0.13
School 5		0.11		0.13
Step 2	0.03		0.01	
Gender		$0.16^{\wedge}$		0.07
Pressure while Learning		-0.07		-0.03
Step 3	$0.02^{\wedge}$		0.03^{*}	
Gender x Pressure while Learning		$-0.27^{\wedge}$		-0.32^{*}
Total R^2	0.27		0.17	

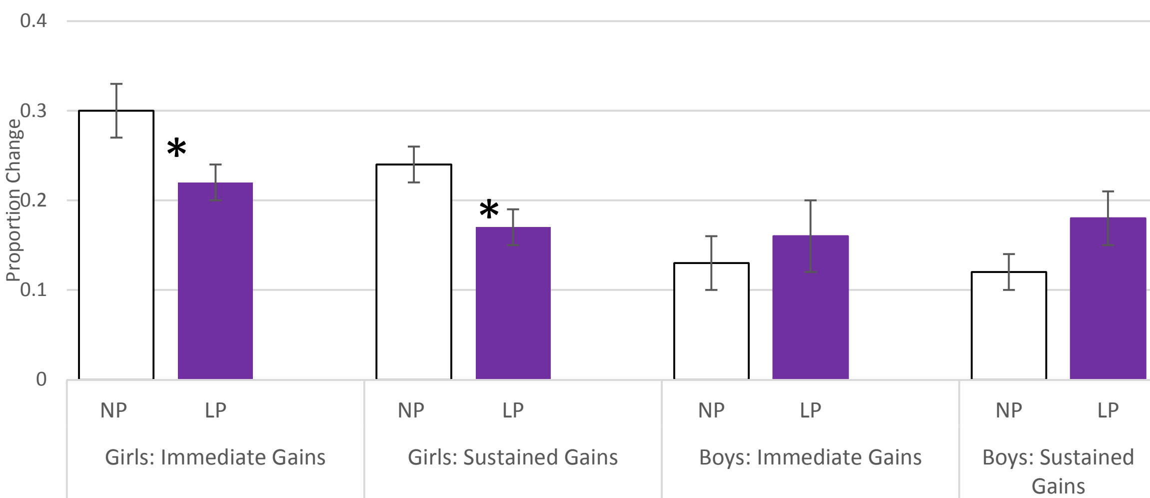
Note. All effect sizes are standardized

^ p <0.10 * p <0.05. ** p <0.01.

Table 3.4 Regression analysis showing main effects and interactions of the learning pressure (LP) study manipulation and student gender in predicting immediate and sustained learning gains

To better understand these interactions, I next examined effects of the LP study manipulation among boys and girls separately. The analyses indicate that, among girls, heightened pressure during instruction predicted smaller learning gains (Table 3.5, Figure 3.2), suggesting that pressure acted more as a distracting threat than as a motivating incentive. Girls assigned to the LP condition had smaller immediate ($\beta_{standardized}$ = -0.26, p =0.04) and sustained ($\beta_{standardized}$ = -0.29, p =0.03) learning gains. In contrast, the pressure manipulation did not harm boys' learning (Table 3.6). Instead, boys who were assigned to the LP condition actually had numerically larger learning gains as compared to boys assigned to the control condition, although these differences were not statistically significant (Figure 3.2).

Figure 3.2. Learning Gains among boys and Girls in the Control (NP) vs. Learning Pressure (LP) Study Conditions



	Immediate Gains		Sustained Gains	
Predictor	**Δ** R^2	*B*	**Δ** R^2	*B*
Step 1	0.28^{**}		0.20^{*}	
Pretest		-0.45^{**}		-0.33^{*}
School 1		-0.34^{*}		$-0.28^{\wedge}$
School 2		-0.10		-0.14
School 3		-0.17		0.01
School 5		0.05		0.14
Step 2	0.06^{*}		0.07^{*}	
Pressure while Learning		-0.26^{*}		-0.29^{*}
Total R^2	0.34		0.28	

Table 3.5. Regression analysis showing impacts of Learning under Pressure among Girls

	Immediate Gains		Sustained Gains	
Predictor	**Δ** R^2	*B*	**Δ** R^2	*B*
Step 1	0.19*		0.15	
Pretest		-0.39		-0.34*
School 1		-0.05		-0.02
School 2		-0.15		-0.06
School 3		-0.28*		-0.27
School 5		0.16		0.10
Step 2				
Pressure while Learning	0.01	0.07	0.02	0.13
Total R^2	0.20		0.17	

^ $p < 0.10$ * $p < 0.05$. ** $p < 0.01$.

Table 3.6. Regression analysis showing impacts of Learning under Pressure among Boys

Gender Differences in impacts of pressure during testing on learning outcomes

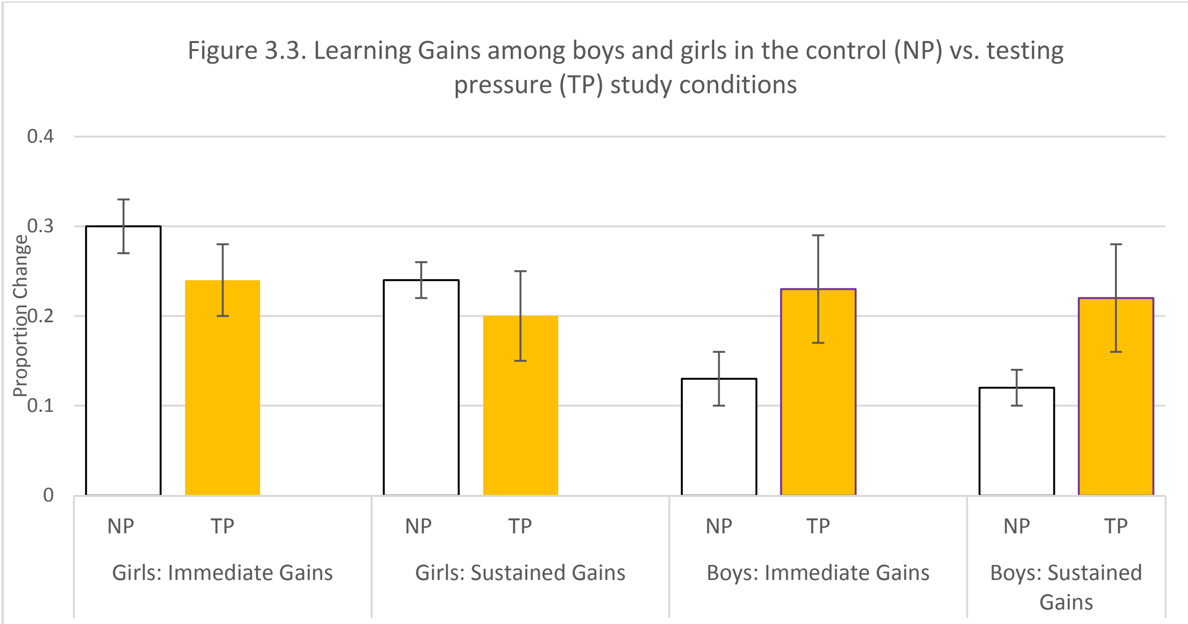
Next, to examine possible gender differences in the role of pressure experienced in testing contexts, I examined main effects and interactions of the Testing Pressure (TP) study manipulation and student gender. As previously, pretest score, along with a dummy variable for school were first entered into the regression (Step 1), as control variables. Main effects (Student gender, TP study manipulation) were added at step 2. Finally, to test the possibility that impacts of heightened pressure during testing differed for boys and girl in this study, a testing pressure * gender interaction term was added to the analysis (Model 3). In contrast to findings for the learning pressure study manipulation, the analysis indicated no main effects of the testing pressure manipulation, or interactions with student gender (Table 3.7). Although the analysis did not indicate a gender * testing pressure interaction, girls assigned to the testing pressure condition had numerically smaller learning gains than did girls assigned to the no pressure condition, while among boys the reverse pattern was seen (Figure 3.3)

	Immediate Gains		**Sustained Gains**	
Predictor	**Δ** R^2	B	**Δ** R^2	B
Step 1	0.21^{***}		0.16^{**}	
Pretest		-0.44^{**}		-0.38^{***}
School 1		-0.23^{*}		-0.16
School 2		-0.09		-0.03
School 3		-0.19		-0.15
School 5		0.05		0.10
Step 2	0.03		0.01	
Gender		0.17^{*}		$0.22^{\wedge}$
Pressure while Testing		0.04		0.16
Step 3	0.12		0.01	
Gender x Pressure while Testing		-0.22		-0.20
Total R^2	0.25		0.19	

Note. All effect sizes are standardized

^ p <0.10 * p <0.05. ** p <0.01.

Table 3.7. Regression analysis showing main effects and interactions of the testing pressure study manipulation and student gender in predicting immediate and sustained learning gains



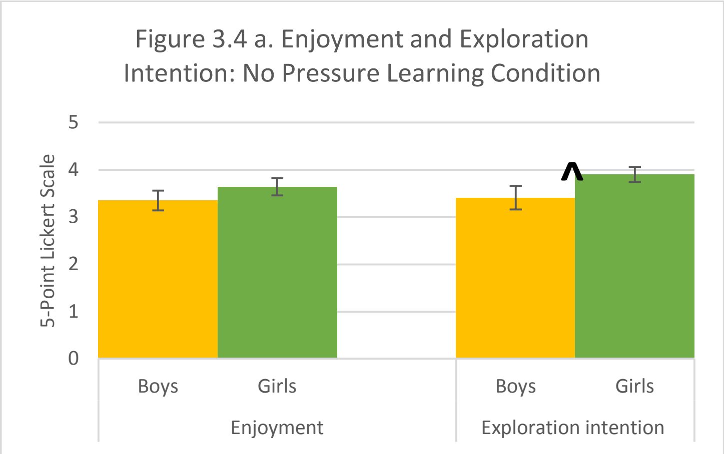
Gender Gaps in engagement in the high vs. low pressure experimental conditions

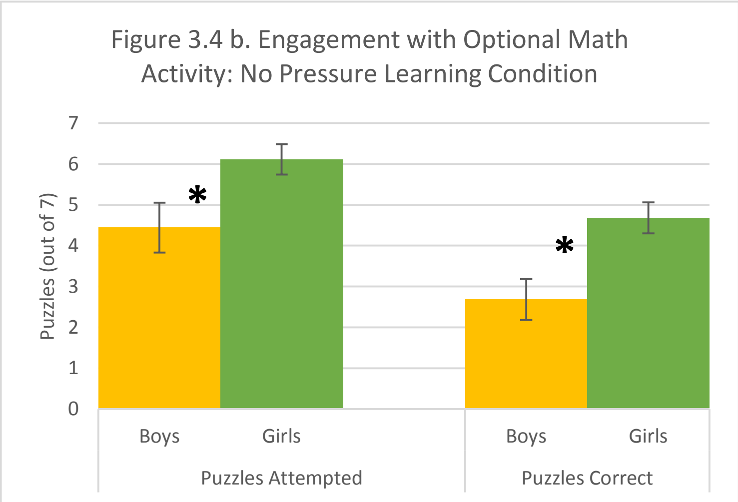
I conducted a series of single linear regressions to test for gender differences in engagement (enjoyment, exploration intention, and likelihood of completing a non-required math activity). Mirroring findings for learning outcomes, in the no pressure condition, girls exhibited higher engagement as compared to their male counterparts (Table 3.8, Figures 3.4a and b). Specifically, in the no pressure learning condition, female students attempted more optional math puzzles ($R^2 = 0.10$, $B = 1.63$, $SE\ B = 0.69$, $t = 2.38$, $p = 0.02$), and completed a greater number of these puzzles successfully ($R^2 = 0.17$, $B = 1.96$, $SE\ B = 0.61$, $t = 3.21$, $p = 0.002$). Additionally, girls in the no pressure condition trended towards reporting greater exploration intention and numerically reported greater enjoyment.

	Exploration Intention			Enjoyment			Puzzles Attempted			Puzzles correct		
	R^2	*B* (SE)	*t*	R^2	*B* *(SE)*	*t*	R^2	*B* (SE)	*t*	R^2	*B* *(SE)*	*t*
Gender	0.05	0.47^ (0.29)	1.66	0.02	0.30 (0.27)	1.10	0.10	1.63* (0.69)	2.38	0.17	1.96** (0.61)	3.21

^ p <0.10 * p <0.05. ** p <0.01.

Table 3.8. *Single regression analysis showing relations between gender and engagement outcomes in the no pressure (NP) learning condition*





I next examined whether gender predicted these same outcomes among students experiencing heightened pressure during learning or testing. Among students who experienced heightened pressure while learning or testing , the gender gap in engagement not only disappeared, but in some cases reversed, mirroring findings on learning outcomes

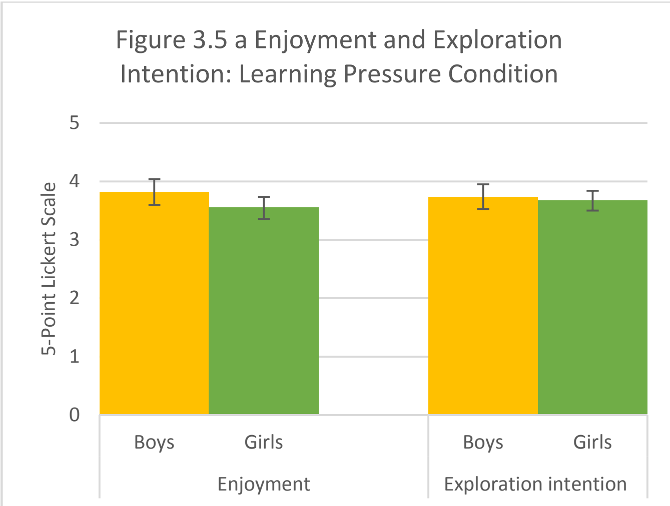
Among students in the Learning Pressure study condition, boys completed more optional math puzzles successfully ($R^2 = 0.06$, $B = -1.11$, $SE\ B = 0.55$, $t = -2.03$, $p = 0.045$), and reported numerically greater enjoyment and exploration intention (Table 3.9, Figures 3.5a and b).

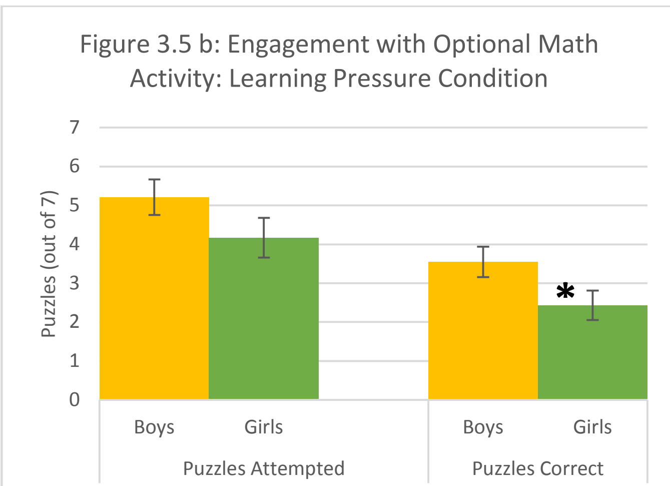
Among students who experienced pressure while testing, boys reported higher enjoyment ($R^2 = 0.07$, $B = -0.58$, $SE\ B = 0.27$, $t = -2.17$, $p = 0.03$; Table 3.10, Figures 3.6 a and b).

	Exploration Intention			Enjoyment			Puzzles Attempted			Puzzles correct		
	R^2	*B* (SE)	*t*	R^2	*B* *(SE)*	*t*	R^2	*B* (SE)	*t*	R^2	*B* *(SE)*	*t*
Gender	0.001	-0.08 (0.28)	-0.27	0.01	-0.27 (0.29)	-0.92	0.04	-1.05 (0.69)	-1.52	0.06	-1.11* (0.55)	-2.03

^ $p < 0.10$ * $p < 0.05$. ** $p < 0.01$.

Table 3.9 *Single regression analysis showing relations between gender and engagement outcomes in the learning pressure learning condition*





	Exploration Intention			Enjoyment			Puzzles Attempted			Puzzles correct		
	R^2	*B* (SE)	*t*	R^2	*B* *(SE)*	*t*	R^2	*B* (SE)	*t*	R^2	*B* *(SE)*	*t*
Gender	0.04	-0.46^ (0.28)	-0.21	0.07	-0.58* (0.27)	-2.17	0.007	0.46 (0.73)	0.73	0.07	1.30 (0.61)	2.12

^ p <0.10 * p <0.05. ** p <0.01.

Table 3.10 *Single regression analysis showing relations between gender and engagement outcomes in the testing pressure learning condition*

Figure 3.6 a Enjoyment and Exploration Intention: Testing Pressure Condition

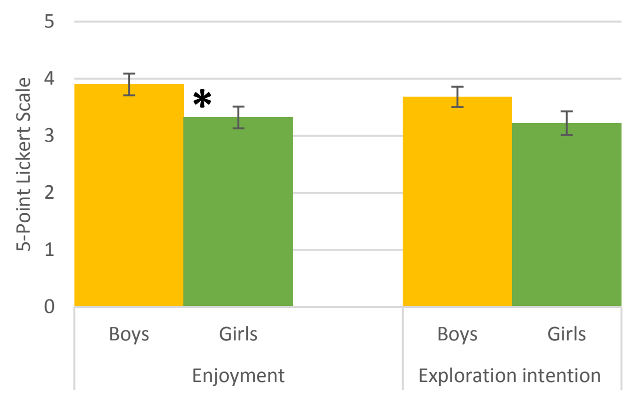
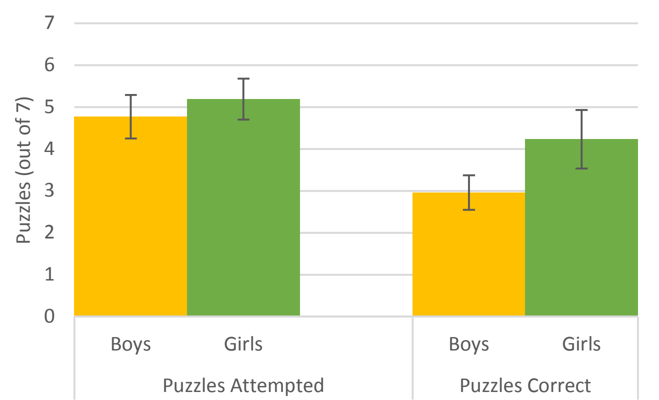


Figure 3.6 b. Engagement with Optional Math Activity: Testing Pressure Condition



Discussion

The findings reported here provide support for the possibility that gender differences in the role of pressure in shaping mathematics learning and engagement may help to explain the seemingly paradoxical ways in which mathematics achievement remains patterned by gender. Among students assigned to the No Pressure study condition, girls not only learned more, but also exhibited higher engagement outcomes. In contrast, gender gaps in learning disappeared in the pressure conditions, with boys and girls showing similar learning gains when pressure was experienced either before or after learning. With the introduction of pressure, gender gaps in engagement actually reversed, with boys who learned under pressure being more likely to engage in optional math activities and those who tested under pressure reporting greater enjoyment.

The reversal of gender gaps across the low pressure versus heightened pressure experimental conditions raises the question as to whether this is due to heightened pressure facilitating boys' mathematics learning and engagement or harming girls' mathematics learning and engagement. Did pressure during learning help boys or harm girls? How about pressure during testing? Answering these questions has important implications for practice because better identifying when pressure helps versus harms learning could help to support learning for all students by allowing educators to leverage the potential for pressure to act as a motivator- while minimizing its potential to act as a distracting threat.

The clearest answer from this experiment is that experiencing pressure during learning was harmful for girls (on average). Girls who experienced pressure while learning had significantly smaller learning gains immediately following the lesson and these differences persisted one week later- even when pressure was no longer heightened. Compounding these direct effects on learning, girls who experienced pressure while testing were less likely to attempt and complete optional math activities. However, trends in the data suggest that the disappearance or reversal of the gender gap in the heightened pressure experimental conditions may also be partially due to pressure boosting boys' learning and engagement outcomes.

In the no pressure condition, girls learned more from the lesson than their male classmates and were also more likely to complete optional math puzzles. Experiencing pressure, however, predicted worse learning, performance, interest and motivation among girls- particularly when the pressure was experienced during learning. This may be because, even in the no pressure condition, without the prospect of receiving an incentive, girls were already engaging effort-fully with the lesson and assessments. Raising the stakes may have resulted in intrusive thoughts and worries that interfered with cognitive engagement while having limited incentivizing effects as girls were already engaged in the no pressure condition. In other words, girls appear to have started out in the optimal part of Yerkes & Dodson's (1908) "inverted-U", and the pressure manipulation may have pushed them into the right side of the U, where too much stress interferes with learning and performance outcomes. This may have been the case particularly if girls were more math anxious or if an increased emphasis on performance invoked concerns among girls that they would be judged stereotypically.

In contrast, among boys, pressure appears to have possibly facilitated better learning and performance outcomes, bringing boys' performance, interest and motivation up to the level at which girls were at in the no pressure condition. Raising the stakes may have motivated boys, who otherwise may have been more likely to disengage during instruction or testing, to put forth increased effort. Or, in other words, boys may have started out on the left side of Yerkes & Dodson's (1908) "inverted-U", where learning and performance suffers due to inadequate task engagement and effort, and the prospect of being rewarded for performance may have pushed them into an optimal level of stress.

The findings reported here suggest the possibility that not feeling pressured enough may contribute to male's relative underachievement on measures of mathematics achievement requiring sustained effort in low-stakes settings (e.g. Easton, Johnson & Sartain, 2017; Gershenson, & Holt, 2015), while experiences of too much pressure, possibly intersecting with mathematics anxiety and gender stereotypes, may contribute to female's relative underachievement in high stakes mathematics settings such as competitions (e.g. Ellison& Swanso (2018), and eventual attrition from math-intensive fields (NSF, 2017). Future research that probes what, exactly, it was about pressure that interfered with girls' learning and engagement outcomes while supporting those of boys could help to identify ways to support mathematics achievement for all students across both high and low stakes settings.

CHAPTER FOUR: STEREOTYPE THREAT AND EVALUATIVE PERFORMACNE PRESSURE

Introduction

Experiments 1 and 2 investigated impacts of stereotype threat and of evaluative performance pressure separately. Findings from this first portion of my dissertation indicate that experiencing either pressure source during initial learning opportunities can impact mathematics learning and engagement- for better and for worse. Results highlight that understanding the role of pressure in shaping mathematics learning requires taking into account student characteristics as well as the pressure source. Student gender and baseline EFs in particular both played important roles in shaping how learning and engagement was impacted by experiencing heightened pressure during conceptual math instruction. Additionally, the impacts of stereotype threat and evaluative performance pressure differed- as did the role of student gender and EFs in moderating effects.

Although pressure sources were examined separately in Experiments 1 and 2, stereotype threat and evaluative performance pressure often co-occur in actual classrooms, with students of color being disproportionately likely to attend schools that are affected by high stakes accountability policies (e.g. Au, 2009; Croft, Roberts & Stenhouse, 2015; Wasserberg & Rottman, 2016; Wasserberg, 2017). These two stressors may interact and even exacerbate each other, as fears of being judged stereotypically can be intensified in the context of evaluative pressure related to high-stakes testing (Wasserberg & Rottman, 2016). For example, in qualitative interviews, African American students attending a high school that had adopted a test-centric curriculum after receiving a failing grade from the state in hopes of avoiding sanctions reported experiencing high levels of both evaluative pressure and stereotype threat (Wasserberg

& Rottman, 2016). In the interviews, the high school students reported that their teachers made constant references to the state tests, and that fear of being negatively stereotyped exacerbated pressures associated with so constant a focus on test performance. The students described feeling that their school was stereotyped as being a “dumb” school, and the students attending it as being “dumb, gang-affiliated troublemakers”, and, in one student’s words, the high-stakes testing context exacerbated the effects of these stereotypes: “Some of them think we’re going to be a D school again next year, or an F school again next year, or we’re going to get worse so what’s the point of working? Then there’s the other kids who are nervous because they feel like the school depends on them to raise the school’s grade.” (Wasserberg & Rottman, 2016).

Although it is clear that many students experience both stereotype threat and evaluative performance pressure in their classrooms, limited research has addressed the effects of these stressors when experienced together. Are effects simply additive, or do they interact? Regarding detrimental effects of pressure, is experiencing two sources of pressure worse than experiencing just one? On the one hand, it is possible that if students are already experiencing distracting and intrusive thoughts associated with one pressure source, adding another pressure source would not further impede learning. But, on the other hand, it could also be the case that experiencing multiple sources of pressure could intensify intrusive thoughts and be more damaging to learning than experiencing a single pressure source.

There are also questions to be answered regarding some of the more positive aspects of pressure when identity threatening and non-identity threatening pressure sources are experienced simultaneously. Does the potential for evaluative performacne pressure to function as a motivator and enhance learning and engagement observed among boys in Experiment 2 extend to students who are also experiencing stereotype threat? Or, do feelings of identity threat block out the potential for pressure to function as a motivating incentive? Alternatively, could the potential for non-identity threatening pressure to act as a motivator actually counteract detrimental effects of stereotype threat?

In Experiment 3, I investigated effects of both sources of pressure on students' mathematics learning and engagement. To do so, I ran a 2 x 2 research design crossing stereotype threat and evaluative performance pressure. The primary focus of Experiment 3 is to better understand how experiencing both stereotype threat and evaluative performance pressure simultaneously impacts children's mathematics learning and engagement.

Additionally, while findings from Experiments 1 and 2 indicated that the impacts of experiencing stereotype threat during learning differed from those of experiencing evaluative performance pressure during learning and that these two stressors should not be treated as one and the same, because data were collected in different contexts, a direct comparison of these two stressors was not possible from the first two studies alone. In addition to allowing for an investigation of the impacts of stereotype threat and evaluative performance pressure experienced concurrently, the design of Experiment 3 allows for a more direct comparison of these two stressors. In this chapter, I examine effects of stereotype threat and evaluative performance pressure on student learning. In addition to testing for main effects, I examine whether these two sources of pressure interact with each other as well as with student gender and EF in predicting learning gains.

In this study, to better understand student characteristics that can support learning, even in high pressure or threatening contexts, I also collected a richer set of information on individual differences and student characteristics. In chapter 5, I explore student characteristics that can promote learning, even in high pressure or threatening contexts. Specifically, I examine relations between students' learning orientations, academic efficacy, attitudes towards novelty and mathematics anxiety and learning gains during the lesson. I test whether these relations differ or are similar in high and low-pressure contexts. The goal of these analyses is to identify student characteristics that support learning generally, and to examine whether these relationships extend to high pressure contexts.

Methods

Design and Procedures

To consider the role of both stereotype threat and evaluative performance pressure in shaping student learning, in Experiment 3, I implemented a 2 x 2 research design crossing stereotype threat and evaluative performance pressure. Participants were randomly assigned to one of four design cells that resulted from crossing two between-subjects factors: stereotype threat before learning (vs. not) and incentivized performance pressure before learning (vs. not; see Figure 4.1).

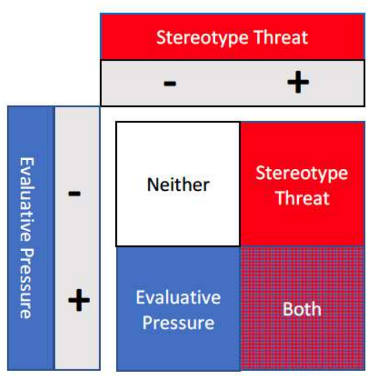


Figure 4.1. Experiment 3 Design

In this study, stereotype threat, evaluative performance pressure, both or neither were invoked before learning using the same pressure manipulations as in the previous experiments. All other study procedures were also identical to those used in the first two experiments, with the exception of several additional measures, all of which were administered during the first study session. These measures were added in order to better understand student characteristics that can support learning, even in high pressure or threatening contexts, and are described and analyzed in Chapter 5.

Participants and Schools

All 5th grade classes at 5 schools in the Chicago area were invited to participate in year 1. One school also participated in Year 2. All schools invited to participate had student populations comprised nearly exclusively of students of color, who are disproportionately burdened by the challenges of contending with stereotype threat in K-12 classrooms. School 1, which participated over 2 study years, is a traditional public elementary school located in a southwest suburb, with a student population comprised of predominantly African American students. School 2 is a charter elementary school located on Chicago's southwest side, serving a predominantly Latinx student population. Schools 3 and 4 are university affiliated charter elementary schools, both located on Chicago's south side and serving predominantly African American Student populations.

A total of ten fifth grade classes (249 students) participated in the study. Thirty-seven students who were absent on one or more study days were excluded from analysis. Ten students who reported their race to be either White (9) or Asian (1) were excluded from the main analyses. Students who reported their race/ethnicity as biracial with at least one negatively academically stereotyped identity were retained in analyses.

The remaining sample included 201 students who were present on all study days and completed all math assessments and learning activities. Self- reported demographics of these participants are shown in Table 4.1 Due to time constraints and/or students getting pulled out of class early, a few students were unable to complete additional measures and surveys. Specifically, 3 students did not complete the PALS survey, 6 students did not complete the

Mathematics Anxiety measure and 12 students did not complete the EF measure. These students were retained in analysis and the missing indicator method was used to handle missing values. Specifically, I created a dummy variable for each of the potentially missing measures (PALs, Math Anxiety, EF). The dummy variable was coded as "1" if the value was missing and "0" if otherwise and was included in the regressions.

	Girl	Boy	Total
Black	73	80	153
Latinx	12	14	26
Biracial*	14	8	22
Total	99	102	201

Table 4.1. *Study 3 Participant Demographics*

* Biracial girls reported their race/ethnicity as: White and African American (4 students), African American and Latino (1 student), African American and Asian (2 students), African American and Other (4 students), Latinx and Other (3 students). Biracial boys reported their race/ethnicity as: White, African American and Latinx (1 student), African American and Latinx (2 students), African American and Other (5 students).

Teacher and Experimenter Demographics

Children's regular math teachers were present in the classroom throughout the study. In five classrooms, the teacher was an African American female. In four classrooms, the teacher was a white female, and in one classroom, the teacher was an African American male. The primary experimenter in all classrooms was either a White or South Asian female.

Results

Analytic Plan

As for the previous experiments, I first report an analysis of pretest performance to ensure that random assignment was successful. I then describe analyses pertaining to my main questions of interest: how experiencing evaluative performance pressure and/or stereotype threat during instruction impacted immediate and sustained learning gains, and whether these pressures interacted with each other or with student gender and EF. I then examine impacts on student's affective engagement with the lesson, including the extent to which students reported that they enjoyed the lesson and desired to learn more about the topic covered in the lesson.

Performance at Pretest

Pretest performance did not differ among students assigned to the four learning conditions or between boys and girls (all p s > 0.15). Pretest performance among boys and girls assigned to each of the learning conditions are shown in Table 4.2.

	Pretest Performance	Immediate Gains	Sustained Gains
Boys	Mean (SE)	Mean (SE)	Mean (SE)
NTP (n=22)	0.14 (0.05)	0.24 (0.06)	0.14 (0.06)
LT (n=27)	0.10 (0.02)	0.10 (0.03)	0.07 (0.03)
LP (n=20)	0.13 (0.04)	0.30 (0.07)	0.28 (0.08)
LTP (n=32)	0.15 (0.05)	0.18 (0.05)	0.13 (0.04)
All (n=101)	0.13 (0.02)	0.19 (0.02)	0.15 (0.03)
Girls	Mean (SE)	Mean (SE)	Mean (SE)
NTP (n=23)	0.05 (0.02)	0.22 (0.05)	0.19 (0.05)
LT (n=24)	0.09 (0.03)	0.30 (0.06)	0.29 (0.06)
LP (n=33)	0.15 (0.04)	0.19 (0.04)	0.12 (0.03)
LTP (n=20)	0.15 (0.05)	0.21 (0.06)	0.16 (0.05)
All (n=100)	0.11 (0.02)	0.23 (0.03)	0.18 (0.02)

Table 4.2. *Mean pretest performance and learning gains among boys and girls assigned to the No Stereotype Threat or Evaluative Pressure (NTP), Learning under Stereotype Threat (LT), Learning under Evaluative Pressure (LP), and Learning under Stereotype Threat and Evaluative Pressure (LTP) Conditions*

Learning under Stereotype threat and/or incentivized performance pressure: Relations between learning condition, student gender and EF

In a series of regressions, I examined main effects and interactions of evaluative performance pressure, stereotype threat, student gender, and EF on immediate and sustained learning gains. Pretest performance and a dummy variable for school were first entered into the model as control variables (Step 1). As discussed in the methods section, I used the missing-indicator method to handle missing EF data, and a dummy variable coded as "1" if EF data was missing, "0" if not was also included in the model at this step. Evaluative performance pressure (binary coded: 0= no evaluative performance pressure; 1=evaluative performance pressure), stereotype threat (binary coded: 0= no stereotype threat; 1= stereotype threat), student gender (binary coded: 0=boy, 1=girl) and EF (entered as a continuous variable with values ranging from 133 to 404) were then added to the model (Step 2, main effects). Next, to examine interactions between gender and learning condition and between EF and learning condition, the following interaction terms were added to the model: evaluative performance pressure x gender, stereotype threat x gender, evaluative performance pressure x EF, and stereotype threat x EF. Finally, in Step 4, I tested for interactions between stereotype threat and evaluative performance pressure. All results from the regression are shown in Table 4.3.

Table 4.3. *Multiple Regression Analyses Predicting Immediate and Sustained Gains in Accuracy From Main Effects and Interactions of Stereotype Threat (ST) and Evaluative Performance Pressure (EPP) with Gender and EF*

	Immediate		**Sustained**	
Predictor	**Δ** R^2	*B*	**Δ** R^2	*B*
Step 1 (Controls)	0.10^{**}		0.07^{*}	
Pretest		-0.20^{**}		-0.15^{*}
School 1		0.10		0.06
School 2		0.18^{*}		0.17^{*}
School 3		0.27^{**}		0.23^{*}
School 4		$0.15^{\wedge}$		0.14
Missing EF Indicator		0.10		0.09
Step 2 (Main Effects)	0.07^{**}		$0.04^{\wedge}$	
Gender		0.07		0.09
EF		0.23^{**}		0.19^{**}
ST		-0.03		0.01
EPP		0.03		-0.004
Step 3 (Interactions)	0.05^{*}		0.06^{**}	
Gender		0.01		0.10
EF		0.26^{*}		0.23^{*}
ST		-0.14		-0.06
IPP		0.22		0.31

ST x Gender		0.29**		0.28*
ST x EF		-0.07		-0.12
EPP x Gender		-0.22^		-0.32**
EPP x EF		-0.04		-0.13
Step 4 (Interaction)	0.00		0.003	
EPP x ST		-0.04		-0.11
Total R^2	0.21		0.18	

^ p <0.10 * p <0.05. ** p <0.01.

Table 4.3. *Multiple Regression Analyses Predicting Immediate and Sustained Gains in Accuracy From Main Effects and Interactions of Stereotype Threat (ST) and Evaluative Performance Pressure (EPP) with Gender and EF* (Continued)

The analysis did not detect main effects of either stereotype threat or evaluative performance on either immediate or sustained learning gains, but revealed significant interactions between learning condition and student gender. Gender and evaluative performance pressure interacted in predicting both immediate ($B_{standardized}$ = - 0.22, p =0.07) and sustained ($B_{standardized}$ =- 0.32, p = 0.009) gains in accuracy. Gender also interacted with stereotype threat to predict immediate ($B_{standardized}$ =0.29, p = 0.009) as well as sustained ($B_{standardized}$ = 0.28, p = 0.01) accuracy gains. The analysis did not detect interactions between incentivized performance pressure and stereotype threat. The analysis showed main effects of student EF; students with higher baseline EFs exhibited larger learning gains, both when measured immediately and at a delay- even when controlling for prior knowledge. EF did not interact with learning condition.

Because the analysis revealed interactions between gender and learning condition- suggesting that effects of stereotype threat and evaluative performance pressure differed for boys and girls-I next examined impacts of evaluative performance pressure and of stereotype threat among boys and girls separately.

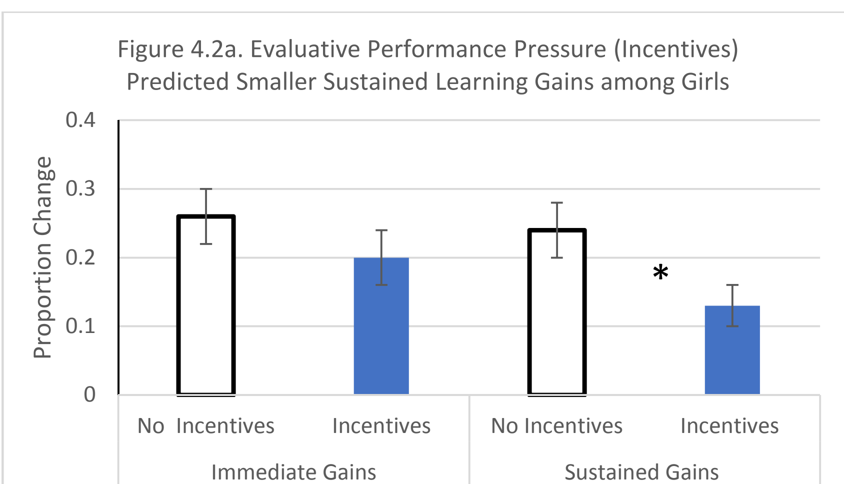
Gender Variations in Main Effects of Evaluative Performance Pressure and Stereotype Threat Manipulations

Evaluative performance pressure. I first examined gender variations in effects of evaluative performance pressure. Evaluative performance (binary coded: 0= no EPP, 1= IPP) was used to predict immediate and sustained gains in accuracy for boys and girls separately, controlling for school, pretest score, and EF. A dummy indicator for whether EF data was missing was also included in the model. All results from the regressions are shown in Tables 4.4a and b. In line with findings in Experiment 2, evaluative performance pressure was generally more harmful for girls and more beneficial for boys (Figures 4.1 a and b). Among girls, the evaluative performance pressure manipulation predicted numerically smaller immediate learning gains and significantly smaller sustained learning gains ($B_{standardized} = -0.21$, $p = 0.04$, Figure 4.2a). In contrast, among boys, the evaluative performance manipulation predicted numerically *larger* immediate gains and marginally significantly *larger* sustained learning gains ($B_{standardized} = 0.17$, $p = 0.08$, Figure 4.2b).

	Immediate		Sustained	
Predictor	**Δ** R^2	*B*	**Δ** R^2	*B*
Step 1	0.17^{*}		$0.13^{\wedge}$	
Pretest		-0.30^{**}		$-0.20^{\wedge}$
School 1		0.09		-0.001
School 2		0.35^{**}		0.27^{*}
School 3		0.28^{*}		$0.21^{\wedge}$
School 4		0.15		0.13
EF		0.15		$0.19^{\wedge}$
Missing EF Ind.		0.05		0.09
Step 2	0.01		0.04^{*}	
EPP		-0.10		-0.21^{*}
Total R^2	0.18		0.17	

^ p <0.10 * p <0.05. ** p <0.01.

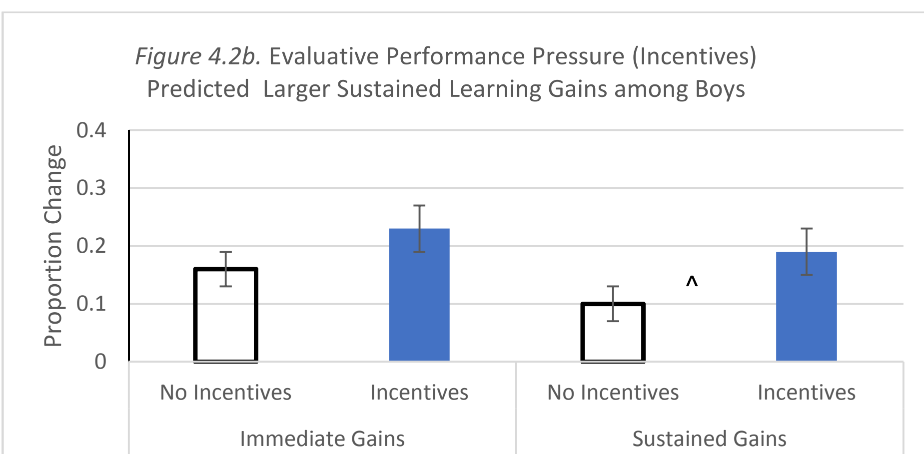
Table 4.4 a. *Multiple regression analysis predicting immediate and sustained learning gains from evaluative performance pressure (EPP) manipulation (girls)*



	Immediate		Sustained	
Predictor	**Δ** R^2	*B*	**Δ** R^2	*B*
Step 1 (Controls)	0.21**		0.11	
Pretest		-0.20*		-0.15
School 1		0.11*		0.12
School 2		0.09		0.14
School 3		0.32*		0.31*
School 4		0.17		0.18
EF		0.33**		0.20*
Missing EF Ind.		0.19*		0.12
Step 2	0.02		0.03$^{\wedge}$	
EPP		0.15		0.17$^{\wedge}$
Total R^2	0.23		0.14	

^ p <0.10 * p <0.05. ** p <0.01.

Table 4.4 b. Multiple regression analysis predicting immediate and sustained learning gains from evaluative performance pressure manipulation (boys)



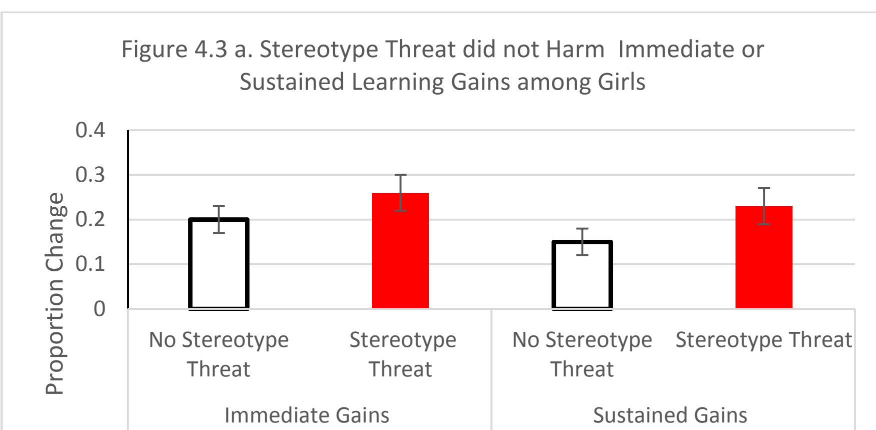
Stereotype threat. I next examined gender variations in effects of the stereotype threat manipulation. Stereotype Threat (binary coded: 0= no ST, 1= ST) was used to predict immediate and sustained gains in accuracy for boys and girls separately. The same controls were used as in the previous analyses: school, pretest score, and EF. A dummy indicator for whether EF data was missing was also included in the model. All results from the regressions are shown in Tables 4.5 a and b. The analysis indicated that, unlike in Experiments 1a and b, girls' learning gains were not harmed by the stereotype threat manipulation. Instead, girls actually exhibited numerically larger sustained learning gains when asked to report on their race before the lesson, although these differences did not approach statistical significance (Figure 4.3a). One possible reason for this contrast with findings in Experiments 1a and b that is considered more fully in the discussion may have been the much greater presence of African American female teachers among participating schools in this study.

Unlike girls, boys who experienced stereotype threat had significantly smaller immediate learning gains ($B_{standardized} = -0.20$ $p = 0.04$) and marginally smaller sustained learning gains ($B_{standardized} = -0.18$, $p = 0.09$). Detrimental effects of stereotype threat were seen among boys regardless of EF level (Figure 4.3b), which also contrasts with findings in Experiments 1a and b. Possible reasons for this difference, which are considered more deeply in the discussion, are differences in students' prior knowledge across studies as well as changes in the broader sociocultural and political context from when experiments 1a and b took place.

	Immediate		**Sustained**	
Predictor	**Δ** R^2	*B*	**Δ** R^2	*B*
Step 1 (Controls)	0.17^{*}		$0.13^{\wedge}$	
Pretest		-0.30^{**}		$-0.20^{\wedge}$
School 1		0.09		-0.001
School 2		0.35^{**}		0.27^{*}
School 3		0.28^{*}		$0.22^{\wedge}$
School 4		0.15		0.13
EF		0.14		$0.19^{\wedge}$
Missing EF ind.		0.05		0.09
Step 2	0.02		$0.04^{\wedge}$	
ST		0.14		$0.19^{\wedge}$
Total R^2	0.19		0.17	

^ p <0.10 * p <0.05. ** p <0.01.

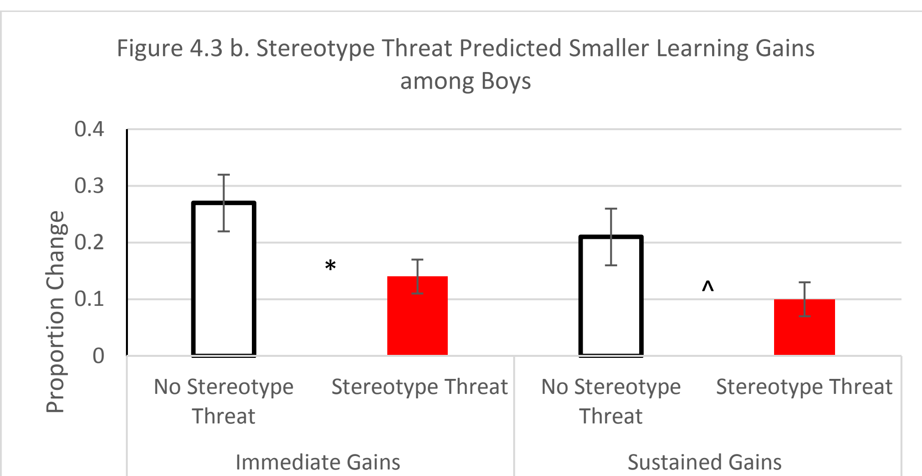
Table 4.5a. *Multiple regression analysis predicting immediate and sustained learning gains from stereotype threat (ST) manipulation (girls)*



	Immediate		Sustained	
Predictor	**Δ** R^2	*B*	**Δ** R^2	*B*
Step 1 (Controls)	0.21^{**}		0.11	
Pretest		-0.20^{*}		-0.15
School 1		0.11		0.12
School 2		0.09		0.15
School 3		0.32^{*}		0.31^{*}
School 4		0.17		0.18
EF		0.33^{**}		0.20^{*}
Missing EF Ind.		0.19^{*}		0.12
Step 2	0.04^{*}		0.02	
ST		-0.19^{*}		-0.15
Total R^2	0.25		0.13	

^ $p < 0.10$ * $p < 0.05$. ** $p < 0.01$.

Table 4.5b. *Multiple regression analysis predicting immediate and sustained learning gains from stereotype threat (ST) manipulation (boys)*



Impacts of Pressure on Engagement Outcomes

While the results discussed above focus on impacts of stereotype threat and evaluative performance pressure on students' learning, it is also important to consider the ways in which experiences of pressure might shape students' affective engagement with mathematics. In the following analyses, I examine effects of the pressure manipulations on the extent to which students reported enjoying the lesson (enjoyment) and feeling motivated to learn more about the topic (exploration intention).

I conducted a series of regressions to test for main effects and interactions between stereotype threat and evaluative performance pressure on students' enjoyment and exploration intention. Stereotype threat and evaluative performance pressure were entered into the model at step 1. The interaction term (stereotype threat x evaluative performance pressure) was added to the model at step 2. All results from the regression are shown in table 4.6. When the interaction term was included in the model, the analysis detected main effects of stereotype threat and of evaluative performance pressure on exploration intention. Students who experienced evaluative performance pressure while learning reported lower exploration intention ($B_{standardized} = -0.25$, $p = 0.01$), and students who experienced stereotype threat while learning also exhibited trends towards lower exploration intention ($B_{standardized} = -0.19$, $p = 0.07$). The analysis also indicated an interaction between the two pressure sources in predicting exploration intention ($B_{standardized} = 0.26$, $p = 0.04$). The analysis did not indicate statistically significant main effects or interactions of stereotype threat or evaluative performance pressure on enjoyment, but students who experienced either pressure source also reported numerically lower enjoyment.

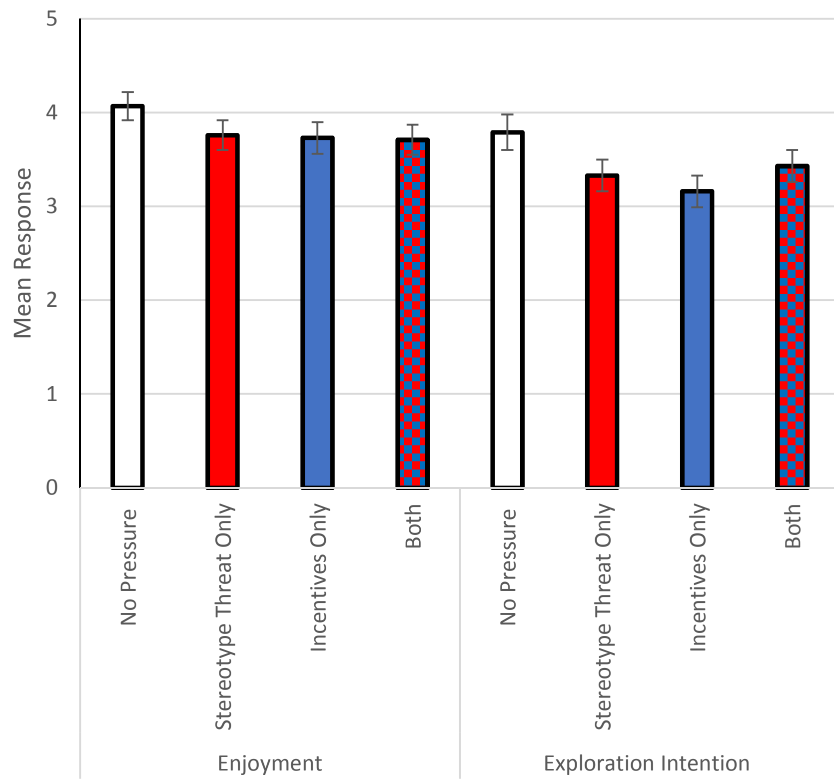
	Enjoyment		**Exploration Intention**	
Predictor	**Δ** R^2	*B*	**Δ** R^2	*B*
Step 1	0.01		0.01	
Stereotype Threat		-0.07		-0.03
Evaluative Performance Pressure		-0.08		-0.10
Step 2	0.004		0.02^{*}	
Stereotype Threat		-0.14		$-0.19^{\wedge}$
Evaluative Performance Pressure		-0.15		-0.25^{*}
Stereotype Threat x Evaluative Performance Pressure		0.11		0.26
Total R^2	0.015		0.03	

^ p <0.10 * p <0.05. ** p <0.01.

Table 4.6. Multiple regression analysis showing impacts of Stereotype Threat and Evaluative Performance Pressure on Enjoyment and Exploration Intention

In the analyses described above, I tested for main effects and interactions of stereotype threat and of evaluative performance pressure. Examining student interest and enjoyment in the four conditions, however, there appeared to possibly be a harmful effect of experiencing any source of pressure on student's enjoyment and exploration intention (see figure 4.4). To explore this possibility, I conducted a regression analysis in which I collapsed across pressure condition, and tested whether being assigned to any of the three pressure conditions predicted differences in interest or enjoyment (Table 4.7). As compared to students in the control group, students in any of the three pressure conditions reported diminished exploration intention ($B_{standardized}$ = -0.16, p=0.02) and possibly lower enjoyment ($B_{standardized}$ = -0.12, p =0.08).

Figure 4.4. Enjoyment and Exploration Intention Among Students in the Four Learning Conditions



	Enjoyment		Exploration Intention	
Predictor	**Δ** R^2	*B*	**Δ** R^2	*B*
Any Threat or Pressure	0.02^	-0.12^	0.03*	-0.16

^ $p < 0.10$ * $p < 0.05$. ** $p < 0.01$.

Table 4.7. *Regression analysis showing impacts of any pressure on enjoyment and exploration intention*

Discussion

Findings in Experiment 3 suggest the existence of complex and non-uniform relations between pressure and academic achievement. One of the most striking findings was the emergence of strong gender differences in the impacts of stereotype threat and of evaluative performance pressure. Bolstering findings reported in the preceding chapters, results from this experiment provide evidence that experiences of pressure during mathematics learning opportunities can have meaningful implications for children's mathematics learning and engagement. While experiences of stereotype threat and evaluative performance pressure in the mathematics classroom can both have meaningful implications for children's cognitive and affective engagement with math, results highlight that identity threatening versus non-identity threatening pressure sources can play different roles in shaping mathematics learning. Results from Experiments 1 and 2 indicated that this was likely the case, and that stereotype threat and evaluative performance pressure should not be treated as one and the same; it is now possible to make this conclusion with greater confidence, as in the current experiment students within the same classroom were randomly assigned to experience the different pressure sources while learning.

Findings from this experiment show the potential for evaluative performance pressure to act as both motivating incentive and distracting threat. Among boys, experiencing evaluative performance pressure appears to have acted as a motivator, boosting learning outcomes- likely through increased effort. Among girls, the same pressure source appears to have acted as a threat, harming learning - likely through increases in intrusive thoughts. These results replicate findings in Experiment 2 and suggest that gender differences in impacts of evaluative performance pressure on learning may play a role in explaining the reversal of gender gaps in mathematics achievement across low vs. high stakes settings.

Results also show the potential for stereotype threat to act as a distracting threat and harm learning. While detrimental effects of stereotype threat were not inevitable, unlike evaluative performance pressure, there was not evidence of stereotype threat actually functioning as a motivating incentive and boosting learning outcomes. Boys of color who were asked to report their race before the lesson had smaller learning gains, while girls' learning was not affected by the stereotype threat manipulation.

There was no evidence that the two different pressure sources interacted. Among boys, these was a main (facilitative) effect of evaluative performance pressure and a main (harmful) effect of stereotype threat regardless of whether or not boys experienced just one or both sources of pressure. Among girls, there was a main (harmful) effect of evaluative pressure that was seen regardless of whether or not girls also experienced stereotype threat. This contrasts with what might be expected based on findings showing that high stakes testing pressures can heighten the salience of stereotypes (e.g. Wasseerberg, 2017), but suggests that stereotype threat and evaluative performance pressure may act through different pathways to shape children's learning.

In Experiment 3, detrimental effects of stereotype threat on learning were seen among boys regardless of EF level, and were not driven solely by high EF students. This contrasts with findings reported in Chapter 2, which detected effects of stereotype threat only among students high in EFs. These cross-study differences are likely real effects, not stemming solely from differences in power across studies. Examination of effect sizes in Experiments 1a and b suggest that stereotype threat truly did not reduce learning gains among low EF students in these studies, and these cross-study differences were not solely because impacts of stereotype threat among low EF students did not reach statistical significance in the earlier studies.

Among low EF students in Experiment 1a, stereotype threat effect sizes for immediate and sustained gains were both extremely small in magnitude (below 0.005) and positive, the opposite direction that would be observed if stereotype threat had harmed learning (see Table 2.3, Figure, 2.2). In Experiment 1b, stereotype threat effects among low EF students were slightly larger in magnitude, but still well below statistically significant effect sizes, and also in the opposite direction as would be expected if stereotype threat had harmed learning (See Table 2.7, Figure 2.3). There are several possibilities for this difference, which I discuss below.

A possible reason that detrimental effects of stereotype threat were seen regardless of student EF level in Experiment 3, but not in Experiments 1a and b, is that participants in Experiment 3 had higher prior knowledge (as measured by pretest performance), particularly as compared with participants in Experiment 1b. Detrimental effects of pressure and cognitive load can be greatest among high EF individuals because when not experiencing cognitive load, these individuals are most able to engage with learning and testing opportunities in cognitively demanding ways, for example engaging with the conceptual underpinnings of comparative mathematics instruction (e.g. Begolli & Richland, 2016; Begolli, Richland, Jaeggi, Lyons, Klostermann & Matlen, 2018; Lyons, Simms, Begoli & Richland, 2018), or using cognitively demanding solution strategies to solve problems (e.g. Maloney, Sattizahn & Beilock, 2014; Maloney, Schaeffer & Beilock, 2013). If low EF students in Experiments 1a and b weren't engaging with the conceptual underpinnings of the lesson and were instead engaging with the lesson on a more superficial level even in the absence of pressure, cognitive load from stereotype threat would be less likely to harm their learning. Among students in the current study, however, higher levels of prior knowledge may have served as a support that enabled students with lower levels of EFs to still engage with the conceptual underpinnings of the lesson when not experiencing intrusive thoughts and cognitive load. This could be a reason that detrimental effects of stereotype threat were seen even among low EF boys in this study.

Changes in the broader sociocultural and political context between Experiments 1a and b and Experiment 3 may also play a role in explaining why detrimental effects of stereotype threat were concentrated among high EF students in the first two experiments, but were seen among boys regardless of EF level in Experiment 3. While children develop awareness of broadly held stereotypes from young ages, the exact age at which children become stereotype aware varies- from child to child to child and depending on the salience of a given stereotype in the child's lived experience. Children who are higher in EFs, for example, may be more cognitively mature and thus may develop awareness of stereotypes at younger ages. Additionally, awareness of highly salient stereotypes tends to develop earlier. Greater salience of discrimination and macroaggression in children's lived experiences may lead children to develop awareness of stereotypes at earlier ages. For example, McKown & Weinstein (2003) found that awareness of broadly held racial stereotypes increased rapidly between age 6 and 10 among all children, but particularly among children of color. At age six, for example, 15 % of Latinx and African

American children in McKown & Weinstein's (2003) study demonstrated awareness of widely held racial stereotypes (compared with 6 % of White and Asian students), while by age ten, the proportion of Latinx and African American children who were stereotype aware had increased to 80% (compared with 63% of White and Asian students).

Data for Experiment 3 was collected during late 2016 and early 2017, a time in which racial stereotypes and tensions were especially salient and visible in the US. Experiments 1a and b took place during 2014 and 2015, a time in which despite still being present, racial tensions and stereotypes were possibly less apparent to young children. It is thus possible that in the earlier context, children lower in EFs may have been less likely to have developed strong stereotype awareness, which may have protected them against experiencing the stereotype threat study manipulation as threatening. In contrast, it is possible that due to the increasing visibility of racial stereotypes in the 2016/2017 context, children who participated in Experiment 3 tended to be stereotype aware by age 10 and 11, regardless of EF level, and were thus more likely to experience the stereotype threat study manipulation as threatening. This is only one possible explanation, and there is no research that I am aware of showing that children's awareness of racial stereotypes increased between these years. However, life course development occurs in "contexts that are shaped by the particular historical moment" (see Spencer & Swanson, 2013) and this possibility therefore should thus not be dismissed.

Another marked difference in findings on impacts of stereotype threat in the current study is that detrimental effects of stereotype threat were only seen among boys, while stereotype threat harmed learning among both boys and girls in Experiments 1a and b. Particularly given increasing salience of racial stereotypes in the broader socio-cultural context during Experiment 3, the finding that stereotype threat did not harm learning among girls is at first surprising. Girls' resilience to the stereotype threat manipulation reminds us that children are embedded in multiple interconnected contexts that include the child's immediate settings as well as more distal influences, such as the sociocultural and political context (Bronfenbrenner, 1979). Supports present in a child's immediate context can help the child to successfully cope with stressors and even mitigate risks from more distal contexts of development, such as heightened salience of racism and racial stereotypes. While none of the students who participated in Experiments 1a and b had African American female math teachers, the majority of participants in the current study did. With the exception of students in one class, even those students who were not themselves taught by an African American female math teacher were still attending schools in which there was an African American female math teacher for their grade. Because students were not randomly assigned to teachers of different races and genders, this conclusion must be considered tentatively, but for girls in Experiment 3, having a female African American math teacher may have served as a support that mitigated against detrimental effects of stereotype threat.

CHAPTER FIVE: STUDENT CHARACTERISTICS THAT SUPPORT LEARNING ACROSS HIGH AND LOW PRESSURE CONTEXTS

Introduction

In chapters two through four, I examined impacts of experiencing heightened pressure during conceptual math instruction among early adolescents and tested whether student gender and baseline EFs moderated effects. In this chapter, I explore the role of additional and more nuanced student characteristics in supporting learning across high and low pressure contexts among the students who participated in Experiment 3. The goal of these analyses is to identify student level protective factors that can promote learning and engagement despite pressure.

The students who participated in all studies learned. In all four experiments and across all study conditions, students showed learning gains, demonstrating greater understanding of ratio on the posttests compared to their pretest understandings. These gains were sustained over at least a week with students continuing to show greater mastery of the material covered in the lesson at the delayed posttest, as compared with their pretest understanding.

This finding should not be considered inconsequential, as students were engaging with a conceptually challenging math lesson in a learning context absent many supports known to facilitate conceptual learning. For instance, although the lesson itself was high quality and incorporated instructional supports to promote relational reasoning, the video-delivered format did not offer affordances for students viewing the lesson to engage in mathematical conversations or to ask questions, both of which can support conceptual math learning (Kazemi & Hintz, 2014; Richland et al., 2016). Additionally, while the lesson's objectives and content were topics within the normal math curriculum for 5th grade children (National Governors

Association Center for Best Practices, 2010), the particular time that students viewed the lesson was often out of sync with their classroom's learning progression. Beyond just the absence of these supports for learning, the experiments introduced additional pressures to challenge learning. Yet, even when contending with both stereotype that and evaluative performance pressure, students learned. What factors contributed to children's successful learning, despite these non-ideal learning contexts?

In the section that follows, I explore student characteristics that promoted learning, despite these challenges. Specifically, I examine relations between student characteristics (including learning orientations, academic efficacy, attitudes towards novelty and mathematics anxiety) and learning and engagement outcomes (students' immediate and sustained learning gains as well as their reported enjoyment and exploration intention following the lesson). I test whether these relations change or remain constant across high and low-pressure contexts. The goal of these analyses is to identify student characteristics that support learning overall, and to examine whether these relationships extend to high pressure contexts.

Learning Orientations

Learning goal orientations, or motivations for engaging in academic behaviors, can be categorized into one of the following three classifications: Mastery Orientation (primary motivation for engaging in academic behavior is developing competence or understanding); Performance-Approach Orientation (primary motivation is demonstrating competence); or Performance-Avoid Orientation (primary motivation is to avoid demonstrating incompetence) (Midgley et. al, 2000). The mastery orientation has been associated with the most positive learning and achievement outcomes, while the performance-avoid orientation has been associated with negative outcomes (Midgley et. al, 2000). Students' learning goal orientations are important predictors of learning and achievement outcomes (e.g. Dweck, 1986; Cury et. al, 2002; Elliot et. al 2005). It is therefore important to understand the extent to which students' goal orientations influence risk and resilience to stressors including stereotype threat and evaluative pressure, when experienced in learning contexts.

Students who endorse an incremental theory of intelligence, or growth mindset, are more likely to possess a mastery goal orientation (e.g. Dweck & Leggett, 1998; Dinger, & Dickhäuser, 2013), and are also less vulnerable to stereotype threat (Aronson, Fried & Good, 2002). Both because of this link and due to the positive learning outcomes generally associated with the mastery orientation, we might expect possessing a mastery goal orientation to be protective against stereotype threat and evaluative pressure. However, a 2011 study by Chalabaev and colleagues that manipulated stereotype threat and goal orientations suggests that the opposite may be true. In this study, inducing a performance avoidance goal orientation buffered the detrimental impact on women's math performance of a gender stereotype threat (Chalabaev et al, 2011). However, it is important to note that this study examined effects of experimentally induced learning goals, as opposed to students' own internalized learning orientations, which could play a different role in supporting learning despite pressure.

In this chapter, I examine relations between children's own learning orientations and their mathematics learning and engagement outcomes following a conceptually challenging mathematics lesson. To help to clarify the relationship between students' goal orientations and vulnerability vs. resilience to stressors including stereotype threat and evaluative pressure, I examine whether introducing pressure changes relationships between children's learning orientations and their mathematics learning and engagement outcomes.

Academic Efficacy

Children's views of themselves as students and beliefs about their academic efficacy are powerful predictors of learning across contexts (Bandura, 1993; Blackwell, Trzesniewski &Dweck, 2007; Eccles & Wigfield, 2002). Academic efficacy could be especially important in high pressure or threatening learning contexts, as children who feel confident their efforts can result in desired outcomes may be more likely to experience heightened pressure as a motivating incentive than as a distracting threat. In this chapter, I examine whether academic efficacy predicts learning and interest outcomes. I then test whether introducing pressure alters these relationships to explore whether academic efficacy may act as a protective factor among students contending with heightened pressure during learning opportunities.

Academic efficacy and mastery goal orientation are both well-known to predict educational achievement (e.g. Ames, 1992; Wolters, 2004; Honicke & Broadbent, 2016). However, how students' academic efficacy and goal orientations shape in-the-moment engagement and learning within a particular learning context is underexplored. The current study examines the role of these student characteristics in shaping in the moment learning and engagement in high and low pressure contexts. I tested: (1) whether student's academic efficacy and goal orientations predicted learning gains from a single high quality, yet challenging, conceptual mathematics lesson, (2) whether these same traits predicted students' enjoyment of the lesson and desire to learn more about the lesson's content, (3) whether these relationships differed across high versus low pressure learning contexts, and (4) whether these relationships differed between boys and girls.

Methods

Participants

Participants were 201 5th grade students. Participating schools and students are described in the preceding chapter and in Table 4.1.

Procedure

Students completed the procedures described in Chapter 4. As noted, these students completed additional measures in order to better understand the role of student level protective factors in promoting learning despite pressure. All of these measures are described below. The analyses reported focus on Mastery Learning Orientation and Academic Efficacy.

Individual Differences Measures

Learning orientations. Students' Learning Orientations, or motivations for engaging in academic behavior, were measured with The Mastery Goal Orientation, Performance-Approach Goal Orientation and Performance-Avoid Goal Orientation subscales from the Patterns of Adaptive Learning (PALS) Instrument (Midgley et al., 2000). The Mastery Goal Orientation subscale (Cronbach's Alpha = 0.85) includes 5 items assessing the extent to which students' reasons for putting forth effort in school include developing their own competence or understanding. The Performance-Approach Goal Orientation subscale (Cronbach's Alpha = 0.89) includes 5 items that assess the extent to which students are motivated by the goal of demonstrating their own competence. The Performance-Avoid Goal Orientation subscale (Cronbach's Alpha = 0.74) includes 4 items that measure the extent to which students are motivated to avoid demonstrating incompetence (Midgley et. al, 2000). Mastery Goal Orientations have been associated with adaptive patterns of learning, while Performance-Avoid Goal Orientations have been associated with maladaptive patterns of learning (Midgley et al., 2000). A Performance-Approach Goal Orientation has been associated with both adaptive and maladaptive patterns of learning (Midgley et al., 2000).

Academic related Perceptions, Beliefs and Strategies. Students' academic related perceptions, beliefs and strategies were assessed using the Academic Efficacy and Avoiding Novelty subscales from the Patterns of Adaptive Learning (PALS) Instrument (Midgley et. al, 2000). The Academic Efficacy (Cronbach's Alpha = 0.78) subscale includes 5 items that gauge students' perceptions of their ability to do school work (Midgley et. al, 2000). The Avoiding Novelty subscale (Cronbach's Alpha =0.78) includes 5 items that gauge students' preference for avoiding unfamiliar school work (Midgley et. al, 2000).

Mathematics Anxiety. The mathematics anxiety measure used in this study was modified from the Single Item Mathematics-Anxiety Scale (SIMA) (Núñez-Peña, Guilera & Suárez-Pellicioni, 2014). SIMA asks adult participants to respond to a single item, "How math anxious are you?" on a Lickert scale. Children in our study were asked to respond to the single item, "Math makes me feel nervous", also on a Lickert Scale. A 2014 validation study of the SIMA shows strong validity and strong test-retest reliability of the single item measure (Núñez-Peña, Guilera & Suárez-Pellicioni, 2014), and scores on the single item mathematics anxiety measure have shown correlations with s-MARS scores of between 0.49 and 0.85 (Ashcraft, 2002).

Results

Analytic Plan

I first report on analyses examining relations between student's academic efficacy, learning orientations, novelty preference, mathematics anxiety and demographic variables including gender and school. I then report on analyses to identify student characteristics that were related to successful mathematics learning and engagement outcomes across learning conditions. These analyses revealed academic efficacy and mastery learning orientation to relate to students' learning and engagement outcomes. I then conducted additional analyses to more closely examine the role of academic efficacy and mastery learning orientation in supporting chilren's in-the moment mathematics learning and engagement in high and low pressure learning contexts. I tested (1) whether these relations differed when students learned under conditions of stereotype threat or evaluative performance pressure and, (2) whether these relations differed between boys and girls.

Relations between student's academic efficacy, learning orientations, novelty preference, mathematics anxiety and demographic variables including gender and school

I examined whether student gender or school predicted learning orientations, academic efficacy novelty preference, or mathematics anxiety. Boys and girls did not differ significantly in learning orientations, academic efficacy, or novelty preference (all ps > 0.1), but girls reported higher mathematics anxiety, F (1, 192) = 6.80, p =0.01.

School did not predict learning orientations, novelty preference, or mathematics anxiety (all ps > 0.2). The analysis, however, did reveal an effect of school on academic efficacy, F (4,194) = 4.186, p <0.001. Post hoc analyses using Bonferroni correction for multiple comparisons indicated that students at both of the university affiliated charter schools had significantly higher academic efficacy, as compared with students at either the other charter school or traditional public school (mean differences from 0.47 to 0.65, significance levels between 0.048 and 0.003). Academic efficacy did not differ between the 2 university-affiliated charter schools, or between the Hispanic serving charter school and traditional public school.

Student Characteristics that Promoted Learning across Conditions

To identify student characteristics that may have supported students' successful mathematics learning and engagement, I examined bivariate correlations between the six student characteristics identified and learning outcomes (immediate gains in accuracy, sustained gains in accuracy), and between the six student characteristics and measures of interest and engagement (enjoyment, exploration intention) (Table 5.1).

Through examination of these bivariate correlations, I identified academic efficacy and mastery orientation as likely supports for learning and engagement, with academic efficacy being most predicted of learning outcomes and mastery orientation most related to engagement. Specifically, results of the Pearson correlation indicated that there were significant positive relations between academic efficacy and both students' immediate learning gains, $r\ (198) = 0.19$, $p = 0.008$, and enjoyment of the lesson, $r\ (198) = 0.24$, $p = 0.001$. There was also a marginally significant positive relation between academic efficacy and sustained learning gains, $r\ (198) = 0.13$, $p = 0.07$. The Pearson correlation also indicated significant positive relations between mastery learning orientation and students' enjoyment of the lesson, $r\ (198) = 0.18$, $p = 0.01$, and desire to explore the topic further $r\ (198) = 0.18$, $p = 0.01$.

	1	2	3	4	5	6	7	8	9	10
1. Immediate Gains	-	0.79^{***}	0.08	-0.02	0.19^{**}	0.09	-0.08	-0.05	-0.08	-0.02
2. Sustained Gains			0.01	-0.03	$0.13^{\wedge}$	0.04	-0.05	-0.10	-0.03	-0.06
3. Enjoyment				0.67^{**}	0.24^{**}	0.18^{**}	0.09	-0.06	-0.09	-0.10
4. Exploration Intention					0.09	0.18^{*}	-0.03	-0.12	-0.10	-0.05
5. Academic Efficacy						0.32^{***}	-0.00	-0.04	-0.33^{***}	-0.35^{***}
6. Mastery Learning Orientation							-0.02	0.08	-0.18^{*}	-0.02
7. Perf. Approach								0.57^{***}	0.23^{**}	0.02
8. Perf. Avoid									0.20^{**}	0.13
9. Avoid Novelty										0.28^{**}
10. Math Anxiety										

^ $p<0.10$ * $p<0.05$. ** $p<0.01$.

Table 5.1. *Bivariate correlations between student characteristics and learning and engagement outcomes*

Risk and Protective Factors for Learning Despite Pressure

Because academic efficacy and mastery learning orientation were identified as both playing roles in supporting student learning and/or engagement during this difficult lesson-despite generally challenging learning contexts, I next examined the role of academic efficacy and mastery orientation more closely. As discussed in the introduction, I hypothesized that both academic efficacy and mastery learning orientation might act as protective factors and could be especially important in supporting learning in high pressure or threatening contexts. To explore this hypothesis, I tested whether these student characteristics interacted with stereotype threat or evaluative performance pressure in predicting mathematics learning and engagement outcomes. Additionally, based on emergent findings showing gender differences in learning and in impacts of the pressure manipulations, I tested whether relations between these student characteristics (academic efficacy and mastery orientation) and learning and engagement outcomes differed between boys and girls.

Academic efficacy and learning gains. I first conducted a series of regressions to test whether the role of academic efficacy in supporting learning remained after controlling for school and pretest performance and to examine whether the role of academic efficacy in supporting learning differed in more versus less pressured contexts or between boys and girls.

Pretest performance and a dummy variable for school were first entered into the model as control variables (Step 1). Academic Efficacy (entered as a continuous variable with values ranging from 1.4 to 5), evaluative performance pressure (binary coded: 0= no evaluative performance pressure; 1=evaluative performance pressure), stereotype threat (binary coded: 0= no stereotype threat; 1= stereotype threat), and student gender were then added to the model (Step 2, main effects). As in previous analyses, I used the missing-indicator method to handle missing data, and a dummy variable coded as "1" if Academic Efficacy data was missing, "0" if not was also included in the model at this step. Next, to test for possible interactions between academic efficacy and learning condition or between academic efficacy and gender, the following interaction terms were added to the model: academic efficacy x evaluative performance pressures, academic efficacy x stereotype threat, academic efficacy x gender (Step 3, interactions). Results of the full regression are shown in Table 5.2.

The results indicate that, even after controlling for school and pretest performance, academic efficacy remained a strong predictor of immediate learning gains ($B_{standardized} = 0.22$, p=0.003) and a possible predictor of sustained learning gains ($B_{standardized}$=0.15, p=0.06). The finding that academic efficacy remained a significant predictor of learning gains even after controlling for pretest performance and school suggests that academic efficacy supported learning during the focal lesson itself and the observed correlation was not simply an artifact of students with greater academic efficacy also having higher prior achievement.

None of the interaction terms included in step 3 were significant, suggesting that the role of academic efficacy in supporting learning extends across high and low pressure contexts and is similar for male and female students.

	Immediate Accuracy Gains		Sustained Accuracy Gains	
Predictor	**Δ** R^2	*B*	**Δ** R^2	*B*
Step 1 (Controls)	0.09*		0.06*	
Pretest		-0.21**		-0.15*
School 1		0.11		0.07
School 2		0.17*		0.17*
School 3		0.27**		0.24**
School 5		0.17^		0.16^
Step 2	0.05*		0.03	
AE		0.22**		0.15^
Missing AE Ind.		0.02		0.03
ST		-0.04		0.00
EPP		0.05		0.01
Gender		0.10		0.11
Step 3	0.01		0.01	
AE x ST		-0.25		0.23
AE x EPP		0.15		0.01
AE x Gender		0.47		-0.48
Total R^2	0.15		0.10	

^ p <0.10 * p <0.05. ** p <0.01.

Table 5.2. Multiple regression analysis showing relations between academic efficacy (AE), gender, stereotype threat (ST), evaluative performance pressure (EPP)and learning outcomes

Academic efficacy and engagement. I next conducted a series of regressions to test whether the role of academic efficacy in supporting engagement differed in more versus less pressured contexts or between boys and girls. Academic Efficacy (entered as a continuous variable with values ranging from 1.4 to 5), evaluative performance pressure (binary coded: 0= no evaluative performance pressure; 1=evaluative performance pressure), stereotype threat (binary coded: 0= no stereotype threat; 1= stereotype threat), and student gender were first entered into the model (Step 1, Main effects). Again, I used the missing-indicator method to handle missing data, and a dummy variable coded as "1" if Academic Efficacy data was missing, "0" if not was also included in the model at this step. Next, to test for possible interactions between academic efficacy and learning condition or between academic efficacy and gender, the following interaction terms were added to the model: academic efficacy x evaluative performance pressure, academic efficacy x stereotype threat, academic efficacy x gender. Results of the full regression are shown in Table 5.3.

As shown, academic efficacy predicted student's enjoyment of the lesson ($B_{standardized}$ =0.23, p=0.001, but did not predict students' desire to explore the topic further. As was the case for the learning outcomes, academic efficacy did not interact with student gender or learning condition to predict engagement outcomes.

	Enjoyment		Exploration Intention	
Predictor	**Δ** R^2	*B*	**Δ** R^2	*B*
Step 1	0.07*		0.02	
Academic Efficacy		0.23**		0.09
Missing Academic Efficacy Indicator		-0.05		0.03
ST		-0.06		-0.02
EPP		-0.07		-0.10
Gender		0.03		0.04
Step 2	0.003		0.003	
Academic Efficacy x ST		-0.21		-0.18
Academic Efficacy x EPP		0.12		-0.06
Academic Efficacy x Gender		0.07		0.18
Total R^2	0.07		0.023	

^ $p < 0.10$ * $p < 0.05$. ** $p < 0.01$.

Table 5.3. *Multiple regression analysis showing relations between academic efficacy, gender, stereotype threat (ST), Evaluative Performance Pressure (EPP) and engagement outcomes*

Mastery learning orientation and learning gains. I conducted a series of regressions to test whether masterly learning orientation predicted learning gains after controlling for school and pretest performance and to examine whether the role of mastery learning orientation in supporting learning differed in more versus less pressured contexts or between boys and girls.

As in the previous analysis, pretest performance and a dummy variable for school were first entered into the model as control variables (Step 1). Mastery learning orientation (entered as a continuous variable with values ranging from 2.6 to 5), evaluative performance pressure (binary coded: 0= no evaluative performance pressure; 1=evaluative performance pressure), stereotype threat (binary coded: 0= no stereotype threat; 1= stereotype threat), and student gender were then added to the model (Step 2, main effects). Again, I used the missing-indicator method to handle missing data, and a dummy variable coded as "1" if Academic Efficacy data was missing, "0" if not was also included in the model at this step. Next, to test for possible interactions between mastery learning orientation and learning condition or between mastery learning orientation and gender, the following interaction terms were added to the model: mastery learning orientation x evaluative performance pressure, mastery learning orientation x stereotype threat, mastery learning orientation x gender (Step 3, interactions). Results of the full regression are shown in Table 5.4. As shown, mastery learning orientation did not predict immediate or sustained learning gains in the full sample. Additionally, mastery learning orientation did not interact with either learning condition or student gender in predicting learning gains.

	Immediate		Sustained	
Predictor	**Δ** R^2	*B*	**Δ** R^2	*B*
Step 1 (Controls)	0.09**		0.07*	
Pretest		-0.21**		-0.15
School 1		0.11		0.07
School 2		0.17*		0.17*
School 3		0.27**		0.24**
School 5		0.17^		0.16^
Step 2	0.02		0.01	
Mastery		0.09		0.02
Missing Mastery		0.04		0.02
ST		-0.06		-0.01
EPP		0.03		0.00
Gender		0.07		0.10
Step 3				
Mastery x ST		-0.10		0.12
Mastery x EPP		-0.14		-0.73
Mastery x Gender		-0.37		0.02
Total R^2	0.11		0.08	

^ $p < 0.10$ * $p < 0.05$. ** $p < 0.01$.

Table 5.4. Multiple Regression Analyses Showing relations between Mastery Learning Orientation, Gender, Stereotype Threat (ST), Evaluative Performance Pressure (EPP) and Learning Outcomes

Mastery learning orientation and engagement. I next conducted a series of regressions to test whether the role of mastery learning goal in supporting engagement differed in more versus less pressured contexts or between boys and girls. Mastery learning orientation (entered as a continuous variable with values ranging from 2.6 to 5), Incentivized performance pressure (binary coded: 0= no incentivized performance pressure; 1=incentivized performance pressure), stereotype threat (binary coded: 0= no stereotype threat; 1= stereotype threat), and student gender were first entered into the model (Step 1, Main effects). Again, I used the missing-indicator method to handle missing data, and a dummy variable coded as "1" if Mastery Learning Orientation data was missing, "0" if not was also included in the model at this step. Next, to test for possible interactions between mastery goal orientation and learning condition or between mastery goal orientation and gender, the following interaction terms were added to the model: mastery goal orientation x incentives, mastery goal orientation x stereotype threat, mastery goal orientation x gender. Results of the full regression are shown in Table 5.5.

As shown, mastery goal orientation predicted both students' enjoyment of the lesson ($B_{standardized}$=0.20, p=0.006) and desire to explore the topic further ($B_{standardized}$=0.19, p=0.009). Mastery learning goal did not interact with either gender or learning condition.

	Enjoyment		Exploration Intention	
Predictor	**Δ** R^2	*B*	**Δ** R^2	*B*
Step 1 (Main Effects)	$0.05^{\wedge}$		$0.05^{\wedge}$	
Mastery		0.20^{**}		0.19^{**}
Missing Mastery		-0.05		0.02
ST		-0.10		-0.05
EPP		-0.09		-0.11
Gender		-0.02		0.01
Step 2 (Interactions)	0.004		0.007	
Mastery x ST		-0.69		-0.91
Mastery x EPP		0.03		0.12
Mastery x Gender		-0.03		-0.01
Total R^2	0.06		0.05	

^ p <0.10 * p <0.05. ** p <0.01.

Table 5.5. *Multiple regression analysis showing relations between mastery learning orientation, gender, stereotype threat (ST), Evaluative Performance Pressure (EPP) and engagement outcomes*

Discussion

Findings from this study suggest that academic efficacy can be a powerful support for learning during challenging math instruction. Students' level of academic efficacy predicted learning gains above and beyond demographic variables and pretest performance- with students higher in in academic efficacy displaying larger learning gains, even after controlling for student's pretest score and school. This finding suggests that academic efficacy supported learning during the focal lesson itself: given the same instructional input, students who were higher in academic efficacy learned more. Students with high academic efficacy also reported enjoying the lesson more, but were no more likely to report a desire to learn more about the topic.

Mastery learning orientation, on the other hand, did not predict learning gains during this lesson, but did predict enjoyment and exploration intention. Students high in mastery learning orientation were more likely to report wanting to learn more about the topic covered in the lesson.

This pattern of results suggests that academic efficacy and mastery goal orientation may contribute to educational achievement outcomes through different pathways. Academic efficacy benefited students through facilitating learning gains during a challenging instructional episode. The benefits of possessing a mastery orientation on the other hand may stem from students with a mastery learning orientation, being more likely to seek out additional learning opportunities that extend and deepen their understanding of concepts introduced within a particular instructional episode.

Academic efficacy did not interact with stereotype threat or incentives to predict engagement or learning outcomes. Instead, the relationship between academic efficacy and learning gains extended across both high and low pressure learning contexts. Likewise, mastery learning orientation did not interact with either of the pressure manipulations to predict engagement outcomes. Having the desire to master new concepts supported students in wanting to learn more about the topic covered in the lesson, even when they had experienced pressure or threat during the lesson itself. These findings suggest that academic efficacy and mastery learning orientations may play an important role in supporting positive academic trajectories, even among students experiencing heightened pressure.

CHAPTER SIX: CONCLUSSIONS

Pressures within the broader socio-cultural context imbue children's everyday mathematics learning experiences. High stakes testing pressures and stereotypes in particular are often salient in the mathematics classroom. Findings from the series of studies reported here indicate that these types of pressures can have meaningful implications for children's mathematics trajectories. In many cases, heightened pressure harmed children's mathematics learning and engagement. Importantly, however, results from these studies highlight that effects of pressure on children's learning are neither homogeneous nor inevitable. Instead, the role of pressure in influencing children's learning and engagement was shaped by the identities and resources students brought to the classroom as well as by the nature of the pressure itself, particularly whether or not identity was implicated. Impacts of pressure were found to depend on the content of the pressure (stereotype threat vs. evaluative pressure) as well as student characteristics, including baseline EFs and gender.

In undertaking this work, I set out to disentangle pressure that acts as a motivator from pressure that acts as a threat in order to better understand: what sources of pressure help vs. harm learning and for which students does pressure help vs. harm learning? The answers to these questions have important theoretical as well as practical implications. Examining the role of pressure content and student level factors in shaping impacts of pressure can help to shed light on mechanisms through which pressure impacts learning and performance. Most importantly, better identification of the situations in which pressure helps vs. harms learning will help to enable educators to avoid harmful effects of pressure while leveraging pressure's motivating qualities in order to improve mathematics learning and engagement for all students.

Findings from each of the experiments reported here show evidence of pressure functioning as a threat. In many cases, pressure experienced during math instruction harmed not only children's learning, but their motivational-affective engagement as well. As predicted given its implications for identity, stereotype threat was especially likely to act as a threat, reducing learning and engagement- particularly among boys and high EF students. While there was evidence that protective factors could in some cases mitigate detrimental effects of stereotype threat, there was no evidence of stereotype threat functioning as a motivator and facilitating learning or engagement. Additionally, findings on impacts of evaluative pressure point to the potential for pressure to function as a threat even when identity is not directly implicated. Although many of the impacts of pressure observed were negative, findings on impacts of evaluative pressure suggest the possibility for heightened pressure to in some cases act not as a threat but as a motivator. Evidence for the role of pressure as motivator was less strong than for the role of pressure as threat, but findings on the impacts of evaluative pressure raise the intriguing possibility that when identity is not implicated, heightened pressure can in some cases actually facilitate children's mathematics learning and engagement.

Evidence of Pressure as Threat across Experiments

Findings from each of the four experiments indicate that a single, short, and relatively subtle experimental manipulation to heighten pressure had detrimental impacts on mathematics learning for at least some groups of students. Identity-threatening pressure was especially likely to function as a distracting threat.

In experiment 1a, a stereotype threat manipulation that increased the salience of racial identity and evaluation before learning had harmful effects on both children's cognitive and motivational-affective mathematics engagement. Specifically, children of color who were asked to report their race and told they would be taking a test following the lesson retained less learning over time- showing smaller learning gains on the delayed posttest, although the threat was no longer present. Additionally, children who experienced stereotype threat while learning enjoyed the lesson less, reported a diminished desire to learn more and were less likely to choose to engage in an optional math activity.

The reduction in sustained learning gains seen among participants in Experiment 1a was greatest among students with high baseline executive function (EF) resources, who otherwise learned most from this lesson. Indeed, the detrimental effects on learning outcomes were driven exclusively by students high in EFs. Findings from Experiment 1b replicated these effects on learning among high EF students. In this experiment as well, when stereotype threat was invoked before the lesson, high EF students retained less learning over time. As in Experiment 1 a, High EF students who experienced stereotype threat while learning showed smaller learning gains at the delayed posttest although the threat was no longer present. In contrast, when stereotype threat was invoked after learning in Experiment 1b, high EF students showed worse performance on the immediate posttest, but detrimental effects of stereotype threat while testing reduced at the delayed posttest when the threat was no longer present.

There are several possible and nonexclusive reasons for the finding that the greatest detrimental effects of stereotype threat on learning were seen among high EF students in Experiments 1a and b. One possibility is that having low EFs was actually protective against *experiencing* the distracting and intrusive thoughts associated with identity threatening pressure. This could have been the case if students lower in attention and inhibitory control were more likely to have been distracted during the study. If these students were looking around the classroom as the study got underway, they may not have been looking at the video screen during the prompt to increase the salience of racial identity and evaluation- and thus were less likely to be affected by the stereotype threat manipulation. Another possibility is that early adolescents lower in EFs may be less meta-cognitively aware, and less likely to have already developed strong awareness of societal stereotypes linking race and mathematics ability. If either of these two explanations were the case, lower EF students may not have actually experienced distracting intrusive thoughts associated with identity threat to the same extent as their higher EF peers.

It is also possible that students lower in EFs did experience intrusive thoughts associated with stereotype threat, but that these thoughts did not have the same detrimental impact on learning as they did among high EF students. This could be the case if even in the low pressure learning condition, low EF students did not engage in the cognitively demanding comparison processes intended in the lesson. If low EF students were more likely to engage with the lesson on a superficial level even in the absence of increased pressure, intrusive thoughts could be less damaging than for high EF students, who would otherwise have engaged deeply with the conceptual underpinnings of the lesson. As EF and pretest performance were often positively related (although this varied across schools), it could also be the case that students lower in EFs were less likely to have the prior mathematical knowledge necessary to engage with and benefit from the lesson-regardless of learning condition.

Findings from Experiment 3 can help to elucidate which of these explanations likely played the greater roles. Experiment 3 also offers evidence of stereotype threat during learning functioning as a distracting threat. In this experiment, detrimental effects of stereotype were seen among both high and low EF students, but only for boys. Specifically, boys who experienced stereotype threat during learning had smaller learning gains. These detrimental effects were seen among boys regardless of EFs. As there were no changes in the experimental protocol between

Experiments 1 and b and Experiment 3 that would have made it less likely that low EF students were simply distracted and not viewing the video screen during the stereotype threat prompt, this explanation for the absence of stereotype threat effects among low EF students in the earlier experiments becomes less likely. It thus seems unlikely that low EF students in Experiments 1a and b did not experience intrusive thoughts simply because they tended to be distracted during the experimental manipulation.

Changes in the broader socio-cultural and political context between Experiments 1a and b and Experiment 3 do support the possibility that low EF students in the first two experiments may have been less likely to experience intrusive thoughts associated with stereotype threat because they were less stereotype aware. The age at which children develop awareness of broadly held stereotypes depends on both child characteristics and the salience of the particular stereotype. Awareness of highly salient stereotypes tends to develop earlier, and children who are more cognitively advanced may also develop awareness of stereotypes at younger ages (McKown & Weinstein, 2003). Experiment 3 took place in late 2016 and early 2017, a time during which explicit racism and racial stereotyping was highly visible in the US, while Experiments 1a and b took place during 2014 and early 2015, a time in which despite still being present, racism and racial stereotypes were more often implicit and may not have been as visible to young children. It is thus possible that in the earlier context, low EF children may have been less likely to have already developed strong awareness of stereotypes linking race and mathematics ability, while in the later context this was less likely to be the case.

As compared with students in the earlier studies, participants in Experiment 3 also had higher prior knowledge, as measured by pretest performance. This difference suggests that it could also have been the case that low EF students in the earlier experiments may have still have experienced the stereotype threat prompt as threatening, but that their learning was less likely to be harmed by these intrusive thoughts. This could have been the case if these students were more likely to engage with the lesson on a superficial level (e.g. memorizing procedures as opposed to drawing connections and comparisons across solution strategies) even in the absence of distracting thoughts. Among low EF students in Experiment 3 on the other hand, it is possible that high prior knowledge served as a support that enabled them to engage with the conceptual underpinnings of the lesson when not experiencing distracting thoughts.

While identity-threatening pressure was especially likely to function as a distracting threat, findings in Experiments 2 and 3 indicate that pressure can also have detrimental consequences, even when identity is not directly implicated. In Experiment 2, raising the stakes by introducing the possibility of earning (or losing) an incentive harmed learning and engagement among girls. Experiment 3 replicated these findings, again showing detrimental effects of the evaluative pressure manipulation only among girls. A possible reason that heightened pressure may have harmed girls' learning is that girls may have already been feeling optimal levels of pressure even without experimentally heightening pressure. Indeed, girls tended to exhibit higher engagement and learning in the No Pressure conditions, suggesting that they may have been starting out in the optimal part of the "inverted-U" relationship between arousal and human cognition, in which learning and performance peak (e.g. Yerkes & Dodson, 1908; Sapolsky, 2015). Experimentally heightening pressure may have created conditions in which too much pressure interfered with girls' learning and performance. Additionally, it is possible that (although the evaluative pressure manipulation did not directly heighten the salience of identity), girls may have in some cases experienced the evaluative pressure manipulation as a threat to identity. This could have been the case if an increased emphasis on performance led girls to worry that their math ability would be judged through the lens of gender stereotypes.

Evidence of Pressure as Motivator across Experiments

Although in many cases pressure harmed mathematics learning and engagement, findings suggest that pressure may also have the potential to function as a motivating incentive, particularly when identity is not threatened. In Experiment 2, boys exhibited lower learning and engagement outcomes as compared to girls in the no pressure condition. However, when pressure was heightened by introducing the possibility of earning (or losing) an incentive, this gender gap disappeared.   Although this was partially due to pressure harming outcomes among girls, trends in the data suggest that pressure may have also facilitated learning and engagement outcomes among boys. Indeed, when pressure was invoked either before or after learning, boys in Experiment 2 had numerically higher immediate and sustained learning gains. If boys were more likely to start out in left hand side of the "inverted-U" where learning and performance suffers due to too little stress (e.g. Yerkes & Dodson, 1908; Sapolsky, 2015) , experimentally heightening pressure may resulted in optimal levels of pressure- the type of context in which students learn best because they are engaged without being overly stressed out.

Findings from Experiment 3 offer additional indications of the potential for pressure to function as a motivating incentive and facilitate learning outcomes. Again in this experiment, when pressure was experimentally heightened by introducing the possibility of earning (or losing) an incentive, boys' learning outcomes improved. In this experiment as well, when pressure was heightened before learning, boys had numerically larger immediate learning gains as well as larger sustained learning gains. Evaluative performance pressure did not interact with the stereotype threat manipulation, suggesting that pressure's potential to function as a motivating incentive can act alongside its role as a distracting threat.

Supporting Learning across High and Low Pressure Contexts

Experiences of stereotype threat and evaluative pressure can pose challenges to mathematics learning. However, findings across experiments indicate that harmful effects of pressure on mathematics learning are not inevitable. Instead, the protective factors children brought to the classroom, as well as the supports present in the classroom, shaped children's learning and engagement when contending with heightened pressure. Importantly, academic efficacy and mastery goal orientation, student characteristics that facilitated learning and engagement in low pressure contexts, continued to do so when pressure was heightened. These student characteristics did not interact with or moderate effects of pressure. Instead, irrespective of learning condition, having a mastery learning orientation and high academic efficacy promoted positive learning and engagement outcomes. Thus, although it does not appear that having a mastery learning orientation or being high in academic efficacy can eliminate harmful effects of pressure, this finding suggests that supporting students to develop academic efficacy and a mastery approach would promote children's mathematics engagement and learning across both high and low pressure contexts.

In addition to student level protective factors, classroom and instructional level supports can promote learning even when students are experiencing pressure. For example, in Experiment 3, evidence of pressure's facilitative role and potential to function as a motivating incentive particularly for boys was seen even among students also experiencing threatening pressure.

Thus, just as experiences of threatening pressure did not diminish the facilitative qualities of academic efficacy and mastery learning orientation, experiences of threatening pressure did not block out the potential for non-identity threatening pressure to act as a motivator.

While none of the supports discussed thus far eliminated detrimental effects of pressure, these findings are encouraging because they suggest that factors that promote learning general do extend to high pressure contexts. Findings in Experiment 3, however, suggest the potential for supports present in the classroom to actually avert detrimental effects of pressure. Specifically, findings in Experiment 3 suggest the possibility that the presence of same gender teachers of color may have supported girls not to experience heightened salience of racial identity as threatening- but perhaps instead as motivating. Future work that more directly tests this hypothesis and examines the particular ways in which teachers with shared identities support learning and engagement in the face of identity threat will contribute to better understanding of ways to support resilience among children from academically stigmatized groups.

It is important to note that even in cases in which detrimental effects of pressure were detected, children learned. In addition to the student and classroom level supports discussed thus far, another support for student learning may have been the lesson itself. The lesson was conceptually challenging yet well supported. As discussed in more detail previously, the lesson incorporated desirable difficulties, such as comparisons across solution strategies including misconceptions, that when well supported promote enduring conceptual understandings (Begolli & Richland, 2016; Bjork, 1994; Richland & McDonough, 2019; Rittle-Hohnson & Star, 20017) . Importantly, the desirable difficulties within the focal lesson were well supported: the teacher used linking gestures to highlight connections between solution strategies, both solution strategies were continuously visibly available to students, and were spatially aligned. Thus, the quality of the lesson itself may have played a role in supporting successful learning despite a low-support/high challenge learning context.

Although testing this hypothesis is beyond the scope of my dissertation, future research that explicitly tests which types of instruction are most effective in high pressure, threatening or low-support learning contexts could help to identify ways to support student learning even in high pressure or low support contexts. A common and understandable response to challenging, chaotic or high pressure learning contexts can be to reduce the rigor of instruction. For example, high stakes testing pressures can lead teachers to focus instruction on ensuring students can execute procedures correctly over ensuring students gain conceptual understanding of the material (Au, 2007, 2009; Flores & Clark, 2009). Additionally, time pressures and worries about students having inadequate prior mathematics knowledge and skills or poor attentional control can lead teachers to believe that drawing connections between multiple solution strategies including misconceptions may be less effective for their students (Begolli et al., 2018). It is also possible that beliefs about which students will benefit from conceptual math instruction may intersect with stereotypes, contributing to students from marginalized groups being less likely to receive high quality conceptual math instruction. Yet, access to conceptual, intrinsically challenging yet well-supported instruction may actually be especially important in high pressure or low support learning contexts.

Broader Impact

In today's rapidly changing and interconnected world, preparing the next generation to participate fully in democracy, as well as for individual economic success, requires that schools not only teach mathematics facts and procedures, but also support students in developing complex mathematical reasoning abilities. Indeed, one of the most important, yet challenging, tasks facing educators is to ensure that all students develop the abstract thinking and critical reasoning skills needed to participate fully in our economy and society. Numerical literacy and fluidity with math concepts and quantitative reasoning skills are especially essential, and have important implications for equity.

In many cases, unequal access to high quality instruction contributes to opportunity and achievement gaps in math. Nearly 30 years ago, the activist Robert Moses argued that students denied access to high quality mathematics instruction "are barred from acquiring the knowledge and skills necessary for participation in an economy driven by rapid technological change" (Moses et. al, 1989), and the importance of deep math knowledge and abstract quantitative reasoning abilities has only grown. Research during the last decades of the 20th century as well as the early years of the 21st century, indicated that low income and minority students were especially likely to receive math instruction that emphasized disconnected concepts, memorizing rules, and arriving at a single correct answer over understanding processes and why an answer is correct) (Anyon, 1980, Atweh, 1998; Ladson). More recent research, however, has not found gaps by SES or race in the amount of procedural vs. conceptual math instruction children receive, and has shown overall increases in coverage in advanced math content during the first decade of the 21st century (e.g Engel, Claessens, Watts, & Farkas, 2016). However, although educational reform efforts and curriculum changes such as the Common Core appear to have increased access to conceptual math instruction by shifting educator attention to the more conceptual, reasoning-based aspects of classroom mathematics, access to high quality math instruction is necessary but not sufficient for supporting children to develop conceptual math understandings. It is also essential that students are able to engage deeply with that instruction: consequently, the contexts in which children learn matter. Differences in the extent to which pressure and identity threat imbue children's everyday mathematics learning experiences can shape mathematics learning and engagement in systemic ways- contributing to mathematics achievement and eventual career trajectories remaining patterned by race and gender in ways that result in math-intensive fields remaining heavily male dominated (National Science Foundation, 2017). This underscores that while increasing access to high quality math instruction is important, it is not enough and that we must also attend to the classroom contexts in which instruction occurs. The extent to which learning is pressured varies greatly across schools, classrooms and individual students, and can have important implications for children's learning. Children spend huge portions of their lives in classrooms, and it is important to ensure that these are identity safe spaces in which children feel motivated and engaged without experiencing threatening pressure.

Findings across experiments indicate that heightened pressure during mathematics learning opportunities can in many cases harm children's mathematics learning and engagement. Detrimental impacts of pressure were neither immense nor universal. However, the finding that a single, short (~ 30 s.), and relatively subtle prompt to heighten pressure before a mathematics lesson often reduced children's learning and engagement is concerning, as for many children experiences of evaluative pressure and stereotype threat in the mathematics classroom are both more frequent and severe. The pressures children feel while learning math are often linked to more salient and stressful consequences than winning or losing a desired incentive (e.g. being retained in grade, failing to get into a desired college, or even concerns that one's school will be closed). Additionally, children of color often experience racial macroaggressions as well as instances of overt racism (e.g. Going & Bianco, 2016) that are far less subtle than the stereotype threat manipulation used in these studies. Future more qualitative, ecologically-embedded and for obvious reasons non- experimental research examining how children experience, make meaning of and cope with these more severe pressures will be important for deepening our understanding of the role of pressure in shaping children's mathematics trajectories. The current work, however, indicates that even subtle threats and pressures may be creating conditions of unequal opportunity for learning and doing math and that these may be contributing to real and persistent gaps in mathematics achievement and participation by race and gender. This highlights that alongside working to improve access to high quality instruction, ensuring that all students are able to benefit fully from this instruction requires that educators work to guard against experiences of pressure and stereotype threat during learning as well as testing opportunities.

REFERENCES

Angrist, J. and Lavy, V. (2009). The effects of high stakes high school achievement awards: Evidence from a randomized trial. *The American Economic Review*, 99(4):1384–1414.

Appel, M. & Kronberger, N. (2012). Stereotypes and the Achievement Gap: Stereotype Threat Prior to Test Taking. *Educational Psychology Review*, *24*(4), 609–635. https://doi-org.proxy.uchicago.edu/10.1007/s10648-012-9200-4

Armenta, B. E. (2010). Stereotype boost and stereotype threat effects: The moderating role of ethnic identification. Cultural Diversity and Ethnic Minority Psychology, 16, 94-98. doi:10.1037/a0017564

Aronson, J. (2009). Reducing Stereotype Threat in classrooms: A review of social-psychological intervention studies on improving the achievement of black students. *Issues & Answers Report*, REL 2009, No. 076. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs.

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. Journal of Experimental Social Psychology, 38, 113–125

Aronson, J., Good, C. (2002). The development and consequences of stereotype vulnerability in adolescents. In F. Pajares and T. Urdan (Eds.), Adolescence and Education, Vol. 2: Academic Motivation of Adolescents. Greenwich, CT: Information Age Publishing

Aronson, J. Lustina, M., Good, C., Keough, K (1999) When white men can't do math: necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology* 35, 29-46

Aronson, J., Steele, C. (2005) Stereotype threat and the fragility of human competence, motivation, and self-concept. In C. Dweck & E. Eliot (Eds) Handbook of Competence and Motivation. New York: Gullford

Ashcraft , M. H. , & Krause , J. A. (2007).Working memory, math performance, and math anxiety . *Psychonomic Bulletin & Review* , 14 , 243 – 248 .

Attali Y, Neeman Z, Schlosser A. (2010). Rise to the challenge or not give a damn: differential performance in high vs. low stakes tests. Unpublished manuscript, Tel-Aviv Univ.

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36, 258-267.

Au, W. (2009). *Unequal by design: High-stakes testing and the standardization of inequality*. New York: Routledge.

Bandura, A. (1993) Perceived Self-Efficacy in Cognitive Development and Functioning. *Educational Psychologist*, 28, 117-148.

Bassok, D., Latham, S., & Rorem, A. (2016). Is Kindergarten the New First Grade? *AERA Open*, *2*(1).

Begolli, K.N., Richland, L.E. (2016). Teaching mathematics by comparison: Analog visibility as a double edged sword. *Educational Psychology, 107,* 194-213.

Begolli, K.N., Richland, L.E., Jaeggi, S. M., Lyons, E., Klostermann, E., Matlen, B. (2018), Executive Function in Relational Reasoning: Incorporating Everyday Classrooms into the Science of Learning. *Thinking and Reasoning*. Pp. 280-313 https://doi.org/10.1080/13546783.2018.1429306

Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science, 17,* 339-343.

Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General*, *133*, 584-600.

Beilock, S. L. & Carr, T. H. (2005). When high-powered people fail: Working memory and "choking under pressure" in math. *Psychological Science*, *16*, 101-105.

Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 33,* 983-998.

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests**.** *Science*, *355*(6323), 389–391.

Bian, L., Leslie, S. J., & Cimpian, A. (2018). Evidence of bias against girls and women in contexts that emphasize intellectual ability**.** *American Psychologist, 73*(9), 1139–1153.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing.* (pp. 185–205). Cambridge, MA: The MIT Press.

Blackwell, L.S., Trzesniewski, K. H.. & C Dweck, C. S. (2007). Implicit Theories of Intelligence Predict Achievement across an Adolescent Transition: A Longitudinal Study and an Intervention. *Child Development*, *78*(1), 246.

Brickenkamp, R., & Zillmer, E. (1998). *The d2 test of attention*. Seattle: Hogrefe & Huber.

Bronfenbrenner, U. (1979). The ecology of human development: Experiments by nature and design. Cambridge, MA: Harvard University Press.

Brown, C., & Leaper, C. (2010). Latina and European American Girls' experiences with academic sexism and their self-concepts in mathematics and science during adolescence. *Sex Roles*. 63: 860-870 .

Burgess, S., Metcalfe, R., & Sadoff, S. (2016). Understanding the response to financial and non-financial incentives in education: Field experimental evidence using high-stakes assessments

Center on Education Policy. (2007). Choices, changes, challenges: Curriculum and Instruction in the NCLB era. Washington, DC: Center on Education Policy (CEP). Retrieved July 20, 2008 from http://www.cep-dc.org/_data/n_0001/resources/live/07107%20 Curriculum-WEB%20FINAL%207%2031%2007.pdf

Chalabaev, A., Major, B., Sarrazin, P., & Cury, F. (2012). When avoiding failure improves performance: Stereotype threat and the impact of performance goals. *Motivation & Emotion*, *36*(2), 130-142. doi:10.1007/s11031-011-9241-x

Chang, H., & Beilock, S. L. (2016). The math anxiety-math performance link and its relation to individual and environmental factors: a review of current behavioral and psychophysiological research. *Current Opinion In Behavioral Sciences*, *10*(Neuroscience of education), 33-38. doi:10.1016/j.cobeha.2016.04.011

Chan, N., & Rosenthal, H. (2014). Working memory moderates stereotype threat effects for adolescents in Hong Kong. *Revue Internationale De Psychologie Sociale-International Review Of Social Psychology*, *27*(3-4), 103-118.

Chen, A., Darst, P. W., & Pangrazi, R. P. (2001). An examination of situational interest and its sources. *British Journal Of Educational Psychology*, *71*, 383.

Chasteen, A., Bhattacharyya, S., Horhota, M., Tam, R., Hasher, L. (2005). How feelings of stereotype threat influence older adults' memory performance. *Experimental Aging Research: An International Journal Devoted to the Scientific Study of the Aging Process* 31 (3)

Chavous, T., M, Rivas-Drake, D., Smalls, C., Griffin, T. & Cogbum, C. (2008). Gender Matters, Too : The Influences of School Racial Discrimination and Racial Identity on Academic Engagement Outcomes Among African American Adolescents. *Developmental Psychology*, (3), 637.

Chen, A., Darst, P. W., & Pangrazi, R. P. (2001). An examination of situational interest and its sources. *British Journal Of Educational Psychology*, *71*, 383.

Cho, S., Holyoak, K. J., & Cannon, T.D. (2007). Analogical reasoning in working memory: Resources shared among relational integration, interference resolution, and maintenance. *Memory & Cognition, 35,* 1445-1455

Christina, R. W., & Bjork, R. A. (1991). Optimizing long-term retention and transfer. In D. Druckman & R. A. Bjork (Eds.), *In the mind's eye: Enhancing human performance* (pp. 23-56). Washington, DC: National Academy Press.

Chestnut, E. K., Lei, R. F., Leslie, S. J., & Cimpian, A. (2018). The myth that only brilliant people are good at math and its implications for diversity. *Education Sciences, 8*(2), 6.

Cimpian, A., & Leslie, S. J. (2017). The brilliance trap: How a misplaced emphasis on genius subtly discourages women and African-Americans from certain academic fields. *Scientific American, 317*, 60–65.

Clark, E. (2005). Review of the d2 Test of Attention. In B. S. Plake, J. C. Impara, & R. A. Spies (Eds.), *The fifteenth mental measurements yearbook*. NE, Lincoln: Buros Institute.

Croft, S. J., Roberts, M. A. & Stenhouse, V. L. (2015). The Perfect Storm of Education Reform: High-Stakes Testing and Teacher Evaluation. *Social Justice*, *42*(1), 70–92.

Croizet, J. C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. Personality and Social Psychology Bulletin, 24 (6): 588–594. doi:10.1177/0146167298246003

Currie, C.E., R.A. Elton, J. Todd and S. Platt: 1997, 'Indicators of socio-economic status for adolescents: the WHO health behaviour in school-aged survey', Health Education Research 12, pp. 385–397.

Cury, F., Elliot, A., Sarrazin, P., Da Fonseca, D., & Rufo, M. (2002). The trichotomous achievement goal model and intrinsic motivation: A sequential mediational analysis. Journal of Experimental Social Psychology, 38, 473–481.

Cvencek, D., Meltzoff, & A.N.. Greenwald, A.G. (2011). Math-Gender Stereotypes in Elementary School Children. *Child Development*, *82*(3), 766.

Davies, L. C., Conner, M., Sedikides, C., & Hutter, R. C. (2016). Math Question Type and Stereotype Threat: Evidence from Educational Settings. *Social Cognition*, *34*(3), 196. doi:10.1521/soco.2016.34.3.196

DeCaro, M. S., Rotar, K. E., Kendra, M. S., & Beilock, S. L. (2010). Diagnosing and alleviating the impact of performance pressure on mathematical problem solving. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *63(8),* 1619-1630.

DeCaro, M. S., Thomas, R. D., Albert, N. B., & Beilock, S. L. (2011). Choking under pressure: Multiple routes to skill failure. *Journal Of Experimental Psychology: General*, *140*(3), 390-406. doi:10.1037/a0023466

Devine, A., Fawcett, & Dowker, A. 2012. “Gender Differences in Mathematics Anxiety and the Relation to Mathematics Performance While Controlling for Test Anxiety.” *Behavioral & Brain Functions* 8 (1): 33–41. doi:10.1186/1744-9081-8-33.

D' hondt, F., Eccles, J. S., Houtte, M., & Stevens, P. A. (2016). Perceived Ethnic Discrimination by Teachers and Ethnic Minority Students' Academic Futility: Can Parents Prepare Their Youth for Better or for Worse? *Journal of Youth and Adolescence* 45 (6): 1075-1089

Diamond, J. B., & Spillane, J. P.(2002). *High Stakes Accountability in Urban Elementary Schools: Challenging or Reproducing Inequality? Institute for Policy Research Working Paper*.

Dinger, F. C., & Dickhäuser, O. (2013). Does implicit theory of intelligence cause achievement goals? Evidence from an experimental study. *International Journal Of Educational Research*, *61*(Achievement Goals and Achievement Goal Orientations in Education), 38-47. doi:10.1016/j.ijer.2013.03.008

Dulude, E., Spillane, J. P., & Dumay, X. (2017). High Stakes Policy and Mandated Curriculum: A Rhetorical Argumentation Analysis to Explore the Social Processes That Shape School Leaders' and Teachers' Strategic Responses. *Educational Policy*, *31*(3), 364–403. https://doi.org/10.1177/0895904815598396

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning And Instruction*, 22, 206-214.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*1040-1048. doi:10.1037/0003-066X.41.10.1040

Easton, J.Q., Johnson, E., & Sartain, L. (2017). *The predictive power of ninth-grade GPA.* Chicago, IL: The University of Chicago Consortium on School Research.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist*, *48*(2), 90–101.

Elkind, D. (1967). Egocentrism in adolescence. *Child Development*, *38*, 1025–1034

Ellison, G. & Swanson, A. (2018). Dynamics of the Gender Gap in High Math Achievement. NBER Working Paper No. 24910 August 2018 JEL No. I20J16

Elliott, L., Charles M., C., Linick, J. L., Grethel, M., Ritchie, A., Gwadz, M. V., & Leonard, N. R. (2015). A multi-method exploratory study of stress, coping, and substance use among high school youth in private schools. doi:10.3389/fpsyg.2015.01028

Elliot, A. J., Shell, M. M., Bouas, H. K., & Maier, M. A. (2005).Achievement goals, performance contingencies, and performanceattainment: An experimental test. *Journal of Educational Psychology*, 97, 630–640.

Ellis, J. M., Rowley, L. L., Nellum, C. J., & Smith, C. D. (2018). From Alienation to Efficacy: An Examination of Racial Identity and Racial Academic Stereotypes among Black Male Adolescents. *Urban Education*, *53*(7), 899–928.

Engel, M., Claessens, A., Watts, T., & Farkas, G. (2016). Mathematics Content Coverage and Student Learning in Kindergarten. *Educational Researcher*, *45*(5), 293–300.

Erikson, E. (1959). Identity and the life cycle: Selected papers. *Psychological Issues,* 1, 1-171.

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal Of School Psychology*, *53*25-44. doi:10.1016/j.jsp.2014.10.002

Flores, B., & Clark, E. (2003). Texas Voices Speak out about High-Stakes Testing: Preservice Teachers, Teachers, and Students. *Current Issues in Education*, *6*

Fogliati, V., Busssy, K. (2013). Stereotype threat reduces motivation to improve: effects of stereotype threat and feedback on women's intentions to improve mathematical ability. *Psychology* of women quarterly. 37(3) 310-324

Foley, A. E., Herts, J. B., Borgonovi, F., Guerriero, S., Levine, S. C., & Beilock, S. L. (2017). The math anxiety-performance link: A global phenomenon. Current Directions in Psychological Science, 26, 52-58.

Forbes, C.E., Duran, K.A., Leitner, J.B., Magerman, A (2015) Stereotype threatening contexts enhance encoding of negative feedback to engender underperformance and anxiety *Social Cognition*, 33 (6): 605-625

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

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. Journal of Educational Psychology, 95 , 393–408.

Gershenson, S., & Holt, S. B. (2015). Gender Gaps in High School Students' Homework Time. *Educational Researcher*, *44*(8), 432-441.

Gottfried, A.E. Marcoulides, G. A., Gottfried, GA., Oliver, P.H., Guerin, D.W. (2007). Multivariate latent change modeling of developmental decline in academic intrinsic math motivation and achievement: Childhood through adolescence. *Longitudinal Modeling of Developmental Processes*, (4), 317.

Goings, R. B., & Bianco, M. (2016). It's Hard To Be Who You Don't See: An Exploration of Black Male High School Students' Perspectives on Becoming Teachers. *Urban Review* (4). 628.

Gottfried, A.E., Fleming, J.S., & Gottfried, A.W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology*, 93, 3–13

Grissom, J. A., & Redding, C. (2016). Discretion and Disproportionality: Explaining the Underrepresentation of High-Achieving Students of Color in Gifted Programs. *AERA Open*, *2*(1):1-25

Gunderson, E. A., Park, D., Maloney, E. A., Beilock, S. L., & Levine, S. C. (2018). Reciprocal relations among motivational frameworks, math anxiety, and math achievement in early elementary school. *Journal of Cognition and Development, 19(1)*, 21-46.

Hardy, L., Mullen, R., & Jones, G. (1996). Knowledge and conscious control of motor actions under stress. *British Journal Of Psychology* 87 (4):621-636.

Harpalani, V., Qadafi, A. K.., Spencer, M. B. (2013). Doll Studies. In P. L. Mason (Ed.), *Encyclopedia of Race and Racism,* 2nd Ed, (pp. 67-70). Detroit: Cengage.

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

Hemphill, F. C.&Vanneman, A. (2011).Achievement Gaps: How Hispanic and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress. *National Center for Education Statistics, Institute of Education Sciences, U.S. Dept. of Education.* Statistical Analysis Report. NCES 2011-459.

Hess, T., Auman, C., Colcombe, S. Rahhal, T. (2003). The impact of stereotype threat on age differences in memory performance. *The Gerontologist.* 58 (1): 3-11

Jalava, N., Joensen, J. S., & Pellas, E. (2015). Grades and rank: Impacts of non-financial incentives on test performance. *Journal Of Economic Behavior And Organization*, *115*(Behavioral Economics of Education), 161-196. doi:10.1016/j.jebo.2014.12.004

Kaba, A. (2005). Progress of African Americans in higher education attainment: The widening gender gap and its current and future implications. Education Policy Analysis Archives, 13(25), 1–34.

Kazemi, E., & Hintz, A. (2014). Intentional talk: how to structure and lead productive mathematical discussions. Portland: Stenhouse Publishers.

Keller, J. (2007). Stereotype threat in classroom settings: the interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *The British Journal Of Educational Psychology*, *77*(Pt 2), 323-338.

Larnell, G. V., Boston, D., & Bragelman, J. (2014). The Stuff of Stereotypes: Toward Unpacking Identity Threats amid African American Students' Learning Experiences. *Journal Of Education*, *194*(1), 49-57.

Legette, K. (2018). School Tracking and Youth Self-Perceptions: Implications for Academic and Racial Identity. *Child Development*, *89*(4), 1311-1327

Leslie, S. J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science, 347*(6219), 262–265.

Leuven, E., Oosterbeek, H., and Klaauw, B. (2010). The effect of financial rewards on students' achievement: Evidence from a randomized experiment4. *Journal of the European Economic Association*, 8(6):1243–1265.

Levitt, S. D., List, J. A., & Sadoff, S. (2016). The Effect of Performance-Based Incentives on Educational Achievement: Evidence from a Randomized Experiment. EconLit with Full Text, EBSCO*host*

Levitt, S., List, J., Neckermann, S & Sadoff, S. (2016). "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance." *American Economic Journal: Economic Policy*, 8 (4): 183-219.

List, J. A., Livingston, J. A., & Neckermann, S. (2018). Do financial incentives crowd out intrinsic motivation to perform on standardized tests? *Economics of Education Review*, *66*, 125–136. https://doi-org.proxy.uchicago.edu/10.1016/j.econedurev.2018.08.002

Luthar, S.S., & Kumar, N.L. (2018). Youth in high-achieving schools: Challenges to mental health and directions for evidence-based interventions. In A. W. Leschied, D. H. Saklofske, and G. L. Flett, *Handbook of School-Based Mental Health Promotion: An Evidence-Informed Framework* (pp. 441-458). New York: Springer

Lyons, E.M., Simms, N., Begolli, K.N., Richland, L.E. (2018). Stereotype Threat Effects on Learning from a Cognitively Demanding Mathematics Lesson. *Cognitive Science*, 42 (2), 678-690

Maloney, E., Sattizahn, J. & Beilock, S.L. (2014). Anxiety and Cognition. *WIREs Cognitive Science*, 5, 403-411.

Maloney, E., Schaeffer, M., & Beilock, S. L. (2013). Mathematics anxiety and stereotype threat: Shared mechanisms, negative consequences, and promising interventions. *Research in Mathematics Education.15:2, 115-128, DOI: 10.1080/14794802.2013.797744*

Mangels, J.A. Good, C. Whiteman, R.C.   Maniscalco, B. &. Dweck, C.S. (2012). Emotion blocks the path to learning under stereotype threat. *Social Cognitive and Affective Neuroscience,* 7 (2): 230-241

Martin, A & Marsh, H. (2005). Motivating boys and motivating girls : Does teacher gender really make a difference? *Educational Pedagogies, Policies and Politics*, (3), 320. R

Martin, A. J., Anderson, J., Bobis, J., Way, J & Vellar, R. (2012). Switching On and Switching Off in Mathematics: An Ecological Study of Future Intent and Disengagement Among Middle School Students. *Journal of Educational Psychology*, (1), 1

Martin, C. L., & Ruble, D. (2004). Children's Search for Gender Cues: Cognitive Perspectives on Gender Development. *Current Directions in Psychological Science*, *13*(2), 67–70. https://doi.org/10.1111/j.0963-7214.2004.00276.x

McGee, E. O. & Bentley, L. (2017) The Troubled Success of Black Women in STEM, *Cognition and Instruction*, 35:4, 265-289, DOI: 10.1080/07370008.2017.1355211

McGee, E., & Spencer, M. B. (2012). Theoretical analysis of resilience and identity: An African American engineer's life story. In E. J. Dixon-Román, & E. W. Gordon (Eds.), *Thinking comprehensively about education: Spaces of educative possibility and their implications for public policy* (pp. 161-178). New York, NY: Routledge.

McKown, C. & Strambler, M.J. (2009). Developmental antecedents and social and academic consequences of stereotype-consciousness in middle childhood. *Child Development, 80, 1643*-1659.

McKown, C., & Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. Child Development, 74, 498-515.

Midgley, C., Maehr, M. L., Hruda, L. A., Anderman, E., Anderman, L., Gheen, M., et al. (2000). Manual for the Patterns of Adaptive Learning Scale. Ann Arbor: University of Michigan

Miller, D. I., Eagly, A. H., & Linn, M. C. (2015). Women's representation in science predicts national gender science stereotypes: evidence from 66 nations. *Journal of Educational Psychology*, 107, 631–644.

Miyake, A., Friedman, N. P. Emerson, M.J., Witzki, A.H., Howerter, A & Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex frontal lobe tasks : A latent variable analysis. *Cognitive Psychology*, (1), 49.

Moran, R. M. R. (2017). The Impact of a High Stakes Teacher Evaluation System: Educator Perspectives on Accountability. *Educational Studies*, *53*(2), 178–193. https://doi-org.proxy.uchicago.edu/10.1080/00131946.2017.1283319

National Mathematics Advisory Panel. Foundations for Success: The Final Report of the National Mathematics Advisory Panel, U.S. Department of Education: Washington, DC, 2008.

Nasir, N. S., Atukpawu, G., O ’Connor, K., Davis, M., Wischnia, S., & Tsang, J. (2009). Wrestling with the legacy of stereotypes: Being African American in math class. In D.B. Martin (Ed.), (pp. 231—248). New York, NY: Routledge.

Nasir, N. S., Snyder, C. R., Shah, N., & Ross, K. M. (2012). Racial storylines and implications for learning. *Human Development*, 55, 28S-301.

Nasir, N. S., McKinney de Royston, M., O'Connor, K., & Wischnia, S. (2017). Knowing About Racial Stereotypes Versus Believing Them. *Urban Education*, *52*(4), 491. doi:10.1177/0042085916672290

Nasir, N.S., & Shah, N. (2011). On Defense: African American Males Making Sense of Racialized Narratives in Mathematics Education. *Journal of African American Males in Education*, *2*(1), 24–45.

National Assessment of Educational Progress (2009). Achievement Gaps
How Black and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress Statistical Analysis Report, *NCES, 2009-455*

National Assessment of Educational Progress (2011). Achievement Gaps: How Hispanic and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress Statistical Analysis Report. *NCES,2011-459*

National Assessment of Educational Progress (2015). School composition and the Black-White achievement gap. *NCES, 2015-018*

National Center for Education Statistics (2016). The Condition of Education: Letter from the Commissioner. http://nces.ed.gov/programs/coe/indicator_cge.asp

National Center for Education Statistics (2017). Status and Trends in the Education of Racial and Ethnic Groups. https://nces.ed.gov/programs/raceindicators/indicator_rbb.asp

National Center for Education Statistics (2018). State Education Reforms, Student Readiness and Progress Through School, Table 5.14. Number of instructional days and hours in the school year, by state: 2018. https://nces.ed.gov/programs/statereform/tab5_14.asp

National Science Foundation, National Center for Science and Engineering Statistics. (2017). *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2017.* Special Report NSF 17-310. Arlington, VA. Available at www.nsf.gov/statistics/wmpd/.

Nesdale, D. and Flesser, D. (2001), Social Identity and the Development of Children's Group Attitudes. Child Development, 72: 506-517. doi:10.1111/1467-8624.00293

Nguyen, H. H., & Ryan, A. M., (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93,*1314-1334.

Phelan, P., Yu, H. C., & Davidson, A. L. (1994). Navigating the psychosocial pressures of adolescence: The voices and experiences of high school youth. *American Educational Research Journal*, *31*(2), 415-447.

Pennington, C.R., Heim, D., Levy, A. R., Larkin, D. T. (2016). Twenty Years of Stereotype Threat Research: A Review of Psychological Mediators. *PLoS ONE* 11(1): e0146487. doi:10.1371/journal.pone.0146487

Picower, B. & Mayorga, E. (2015). "What's Race Got To Do With It?: How Current School Reform Policy Maintains Racial And Economic Inequality". *What's Race Got To Do With It?: How Current School Reform Policy Maintains Racial And Economic Inequality*. Volume 2, https://works.swarthmore.edu/fac-education/86

Plucker, J. A., Burroughs, N., Song, R., & Indiana University, C. for E. and E. P. (2010). Mind the (Other) Gap! The Growing Excellence Gap in K-12 Education. *Center for Evaluation and Education Policy, Indiana University*

Rahman, T., Fox, M. A., Ikoma, S., Gray, L., National Center for Education Statistics (ED), American Institutes for Research (AIR), … Sanametrix. (2017). *Certification Status and Experience of U.S. Public School Teachers: Variations across Student Subgroups. NCES 2017-056*. *National Center for Education Statistics*. National Center for Education Statistics.

Ramirez, G., McDonough, I. M., & Jin, L. (2017). Classroom stress promotes motivated forgetting of mathematics knowledge. *Journal of Educational Psychology*, *109*(6), 812-825. doi:10.1037/edu0000170

Reardon, S. (2008). Differential growth in the Black-White achievement gap during elementary school among initially high- and low-scoring students. Institute for Research on Education Policy & Practice Working Paper 2008-7.

Reardon, S.F., Fahle, E.M., Kalogrides, D., Podolsky, A., & Zárate, R.C. (2018). Gender Achievement Gaps in U.S. School Districts (CEPA Working Paper No.18-13). Retrieved from StanfordCenter for Education Policy Analysis: http://cepa.stanford.edu/wp18-13

Regner, I., Smelding, A. , Gimmig, D., Thinus-Blanc, C., Monteil, J. Huguet, P. (2010). Individual differences in working memory moderate stereotype threat effects. *Psychological Science* 21(11): 1646-1648

Richland, L. E., Begolli, K.N., Simms, N., Frausel, R.R. Lyons, E. A. (2016). Supporting mathematical discussions: The roles of comparison and cognitive load, *Educational Psychology Review*, *29*(1), 41-53. *DOI 10.1007/s10648-016-9382-2*

Richland, L.E., Burchinal, M. (2012). Early executive function predicts reasoning development. *Psychological Science, 24*(1) 87-92. doi: 10.1177/0956797612450883

Richland, L.E., & McDonough, I. M. (2010). Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology*, *35*, 28-43.

Richland, L.E., Morrison, R.G., & Holyoak, K.J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Experimental Child Psychology*, *94*(3), 249–273. doi: 10.1016.j.jecp.2006.002.

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99,*561–574.

Rittle-Johnson, B., & Star, J. R. (2009). Compared to what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology, 101,* 529–544.

Rittle-Johnson, B., Star, J. & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology*, 101 (4), 836-852.

Roediger, H.L, &. Karpicke. J. D. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, *17*(3), 249.

Ryan, A. M., & Nguyen, H.-H. D. (2017). Publication bias and stereotype threat research: A reply to Zigerell. *Journal of Applied Psychology*, *102*(8), 1169–1177. https://doi-org.proxy.uchicago.edu/10.1037/apl0000242.supp (Supplemental)

Rydell, R. J., & Boucher, K. L. (2010). Capitalizing on multiple social identities to prevent stereotype threat: The moderating role of self-esteem. *Personality and Social Psychology Bulletin,* 36, 239-250. doi:10.1177/0146167209355062

Rydell, R. J., & Boucher, K. L. (2017). Stereotype threat and learning. In J. M. Olson (Ed.), *Advances in experimental social psychology.* (Vol. 56, pp. 81–129). San Diego, CA: Elsevier Academic Press

Rydell, R. J., Rydell, M. T., & Boucher, K. L. (2010). The effect of negative performance stereotypes on learning. *Journal of Personality and Social Psychology, 99,* 883-896.

Rydell, R. J., Shiffrin, R., Boucher, K. L., Van Loo, K., & Rydell, M. T. (2010). Stereotype threat prevents perceptual learning. *Proceedings of the National Academy of Sciences (USA), 107,* 14042-14047.

Sapolsky, R. M. (2015). Stress and the brain: individual variability and the inverted-U. *Nature Neuroscience*, *18*(10), 1344–1346.

Sattizahn, J.R., Moser, J.S., & Beilock, S.L. (2016) A Closer Look at Who "Chokes Under Pressure". *Journal of Applied REsearch in Memory and Cognition,* 5(4), 470-477.

Saunders, J., Davis, L., Williams, T., & Williams, J. (2004). Gender differences in self-perceptions and academic outcomes: A study of African American high school students. Journal of Youth and Adolescence,33(1), 81–90.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. Journal of Experimental Social Psychology, 38, 194-201. doi:10.1006/jesp.2001.1500

Schmader, T., & Beilock, S.L. (2012). Mechanisms: An integration of processes that underlie stereotype threat. In T. Schmader & M. Inzlicht (Eds.), *Stereotype Threat: Theory, Process, and Application.* Oxford University Press.

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. *Psychology Review* (2):336-56. doi: 10.1037/0033-295X.115.2.336.

Sedek, G., Krejtz, I., Rydzewska, K., Kaczan, R., & Rycielski, P. (2016). Three functional aspects of working memory as strong predictors of early school achievements: The review and illustrative evidence. *Polish Psychological Bulletin*, *47*(1), 103. doi:10.1515/ppb-2016-0011

Shapiro, J. (2012) The role of stereotype threat in undermining girls and women's performance and interest in STEM fields. Sex Roles. 66, 3: 175-183

Smith, J. L., Sansone, C., & White, P. H. (2007). The stereotyped task engagement process: The role of interest and achievement motivation. Journal of Educational Psychology, 99, 99–114. doi:10.1037/0022- 0663.99.1.99.

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 (1):4-28

Spencer, M. B. (1984). Black Children's Race Awareness, Racial Attitudes, and Self-concept: A Reinterpretation. *Journal of Child Psychology Psychiatry* 25: 433-441.

Spencer, M. B. (1985). Cultural Cognition and Social Cognition as Identity Factors in Black Children's Personal-Social Growth. In Beginnings: Social and Affective Development of Black Children, edited by Margaret Beale Spencer; Geraldine Kearse Brookins; Walter R. Allen, 215-230. Erlbaum Hillsdale NJ.

Spencer, M. B. (2006). Phenomenology and ecological systems theory: Development of diverse groups. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology, vol. 1: Theoretical models of human development, 6th ed.* (pp. 829-893). New York: Wiley Publishers.

Spencer, M. B. (2008). Phenomenology and ecological systems theory: Development of diverse groups. In W. Damon and R. M. Lerner (Eds.), *Child and adolescent development: An advanced course* (pp. 696-735). New York: Wiley Publishers.

Spencer, M.B., & Swanson, D. (2013). Opportunities and challenges to the development of healthy
children and youth living in diverse communities. *Development & Psychopathology, 25*(4pt2), 1551-1566.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797–811.

Stetser, M., and Stillwell, R. (2014). Public High School Four-Year On-Time Graduation Rates and Event Dropout Rates: School Years 2010–11 and 2011–12. First Look (NCES 2014-391). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from http://nces.ed.gov/pubsearch.

Suinn, R. M., Taylor, S., & Edwards, R. W. (1988). Suinn Mathematics Anxiety Rating Scale for Elementary School Students (MARS-E): Psychometric and Normative Data. *Educational & Psychological Measurement*, *48*(4), 979. doi:10.1177/0013164488484013

Sullivan, P.A.,Tobias, S. & McDonough, A. (2006). Perhaps the decision of some students not to engage in learning mathematics in school is deliberate. *Educational Studies in Mathematics* 19 (2):81-99

Swanson, D., Cunningham, M., Spencer, M. B. (2005). Black males' structural conditions, achievement patterns, normative needs, and "opportunities." In O. Fashola (Ed.), *Educating African American males: voices from the field* (pp. 229-254), Corwin Press: Thousand Oaks, CA.

Swanson, D. P., Spencer, M. B., & Petersen, A. (1998). Identity formation in adolescence. In K. Borman & B. Schneider (Eds.), *The adolescent years: Social influences and educational challenges: Ninety-seventh Yearbook of the National Society for the Study of Education, Part I.,* (pp. 18-41). Chicago: The National Society for the Study of Education.

Tapia, M., & Marsh, G. E. (2000). Attitudes toward mathematics instrument: An investigation with middle school students. Bowling Green: Paper presented at the Annual Meeting of the Mid-South Educational Research Association.

Taylor, V.J., Walton, G. M. (2011). Stereotype threat undermines academic learning. *Personality and Social Psychology Bulletin*, 37: 1055-1067

Tempel, T. & Neumann, R. (2014). Stereotype threat, test anxiety, and mathematics performance. *Social Psychology of Education* 17: 491-501

Triplett, C. F., & Barksdale, M. A. (2005). Third Through Sixth Graders' Perceptions of High-Stakes Testing. *Journal Of Literacy Research*, *37*(2), 237-260. doi:10.1207/s15548430jlr3702_5

Vanneman, A., Hamilton, L., Anderson, J.B., & Rahman, T. (2009). How Black and White students in public schools perform in mathematics and reading on the National Assessment of Educational Progress. *National Center for Education Statistics, Institute of Education Sciences, U.S. Dept. of Education.* Statistical Analysis Report. NCES 2009-455

Velez, G., & Spencer, M. B. (2018). Phenomenology and intersectionality: Using PVEST as a frame for adolescent identity formation amid intersecting ecological systems of inequality. In C. E. Santos & R. B. Toomey (Eds.), Envisioning the Integration of an Intersectional Lens in Developmental Science. New Directions for Child and Adolescent Development, 161, 75–90

Waltz, J. A., Lau, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. Memory & Cognition, 28, 1205–1212.

Wasserberg, M.J. & Rottman, A. (2016). Urban High School Students' Perspectives on Test-Centered Curriculum. *American Secondary Education*, *44*(3), 56-71.

Wasserberg, M. J. (2014). Stereotype Threat Effects on African American Children in an Urban Elementary School. *Journal Of Experimental Education*,*82*(4), 502-517.

Wasserberg, M. J. (2017). High-Achieving African American Elementary Students' Perspectives on Standardized Testing and Stereotypes. *Journal Of Negro Education*, *86*(1), 40-51

Watson, C. E., Johanson, M., & Dankiw, J. (2014). Effects of High-Stakes Testing on Third through Fifth Grade Students: Student Voices and Concerns for Educational Leaders. *E Journal Of Organizational Learning & Leadership*, *12*(1), 1-11.

Wigfield, A., Eccles, J.S., Schiefele, U., Roeser, R.W., & Davis-Kean, P. (2006). Development of achievement motivation. In N. Eisenberg (Ed.), Handbook of child psychology:Vol. 3. Social, emotional, and personality development (6th ed., pp. 933–1002). New York: Wiley

Wright, S. B., Matlen, B. J., Baym, C. L., Ferrer, E., and Bunge, S. A. (2008). Neural correlates of fluid reasoning in children and adults. Front. Hum. Neurosci. 1:8.

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.

Young, J. L. Young, J. R., & Capraro, M. M. (2017). Black Girls' Achievement in Middle Grades Mathematics: How Can Socializing Agents Help? *Clearing House*, 90 (3), 70-76

Zigerell, L. J. (2017). Potential publication bias in the stereotype threat literature: Comment on Nguyen and Ryan (2008). Journal of Applied Psychology, 102, 1159–1168. http://dx.doi.org/10.1037/apl0000188

APPENDICES

Appendix A: Assessment Items

1) Andre and Jasmin are both mixing paints. Andre's paint has 4 cups of blue paint and 1 cup of white paint. Jasmin's paint has 5 cups of blue paint and 3 cups of white paint.

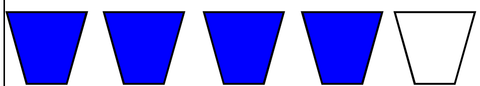
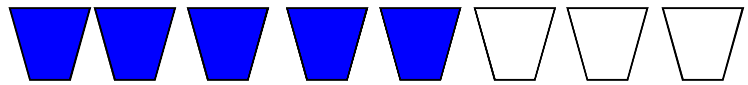
Andre's Paint	Jasmin's Paint
	

Figure A.1 Paints assessment item

Whose paint will be ***darker blue***?

Please show all your work.

Whose paint will be darker blue? ____________

2) Jamal and Sam played video games at Dave and Buster's and then went to turn in their tickets for prizes. For every game they won, they got 1 ticket. Jamal played 27 games and won 11 tickets. Sam played 9 games and won 3 tickets. Who won the bigger portion of games he played?
Please show all your work.

Who won the bigger portion of games played? ____________

3) Ken and Yoko shot several free throws in their basketball game. The result of their shooting is shown in the table. Who made the bigger portion of shots he took?

	Shots Made	Total Shots Tried
Ken	12	20
Yoko	16	25

Table A. 1 Free throws assessment item 1

Please show all your work.

Who made the bigger portion of shots? ___________

How do you know?

a) I compared Ken's shots made with Yoko's shots made
b) I compared Ken's shots tried with Yoko's shots tried
c) I compared Ken's shots made to shots tried with Yoko's shots made to shots tried
d) I compared Ken's shots missed with Yoko's shots missed

4) One of the judges recorded all the missed shots with check marks. Which player had ***more*** check marks?

Please show all your work.

Which player had more checkmarks? ___________

How do you know?

a) I compared Ken's shots made with Yoko's shots made
b) I compared Ken's shots tried with Yoko's shots tried
c) I compared Ken's shots made to shots tried with Yoko's shots made to shots tried
d) I compared Ken's shots missed with Yoko's shots missed

5) What is the definition of "ratio"?

6) Shani and Keisha have both set up lemonade stands. Shani's lemonade recipe uses 2 cups of lemon juice and 1 cup of water. Keisha's lemonade recipe uses 3 cups of lemon juice and 2 cups of water.

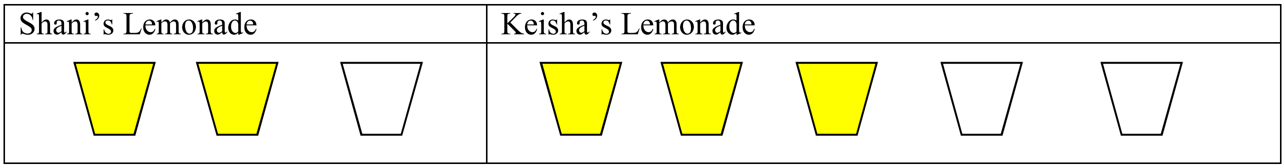


Figure A 2. Lemonade assessment item

Whose lemonade tastes ***more*** "lemony"?

Show all your work.

Whose lemonade tastes more "lemony"? ____________

7) After school, Jenna and Nia played in a free-throw tournament. Their results are in the table below. Who made the bigger portion of shots she took?

	Shots Made	Total Shots Tried
Jenna	8	15
Nia	12	20

Table A2. Free throws assessment item 2

Please show all your work.

Who made the bigger portion of shots? ___________

How do you know?

a) I compared Jenna's shots made with Nia's shots made
b) I compared Jenna's shots tried with Nia's shots tried
c) I compared Jenna's shots made to shots tried with Nia's shots made to shots tried
d) I compared Jenna's shots missed with Nia's shots missed

For the next five questions, please circle your response.

8) Shawna and Deonte each tossed coins and recorded how many times they guessed heads or tails correctly. Shawna tossed 7 coins and guessed 4 correctly. Deonte tossed 21 coins and guessed 12 correctly.
Which of the following strategies will tell us who was better at guessing?

a) *Find the least common multiple for 7 and 21*
b) *Find the least common multiple for 4 and 7*
c) *Both are correct*
d) *Neither are correct*

9) Shawna and Deonte kept track of incorrect guesses with frowny faces ☹.
Which of the following strategies will tell us who had ***more*** frowny faces ☹?

a) *Find the least common multiple for 21 and 7*
b) *Subtract 7 – 4 and 21 – 12*
c) *Both are correct*
d) *Neither are correct*

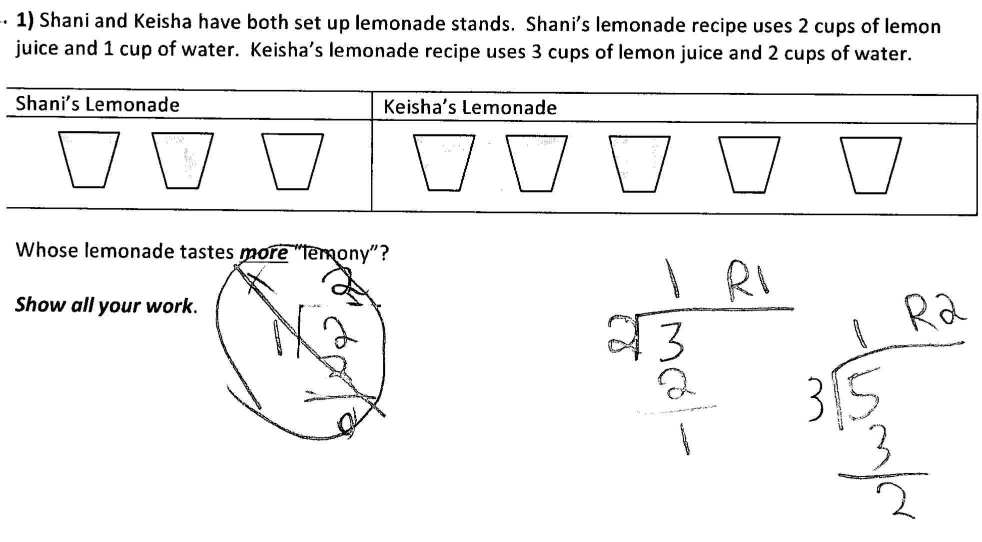
10) A weather channel in Chicago and a weather channel in Miami both tried to predict all the rainy days last month. The Chicago weather channel correctly predicted 16 rainy days out of 20 rainy days total. The Miami weather channel correctly predicted 8 rainy days out of 10 total rainy days.
Which of the following strategies will tell us which channel is better at predicting rain?

a) *Subtract 20 – 16 and 10 – 8*
b) *Find the least common multiple for 20 and 10*
c) *Both are correct*
d) *Neither are correct*

Appendix B: Coding Manual

Coding Manual

1. **Division**
   a. Long Division
   b. Division using the division sign
   c. Example
   Figure B 1. Division example
   d.

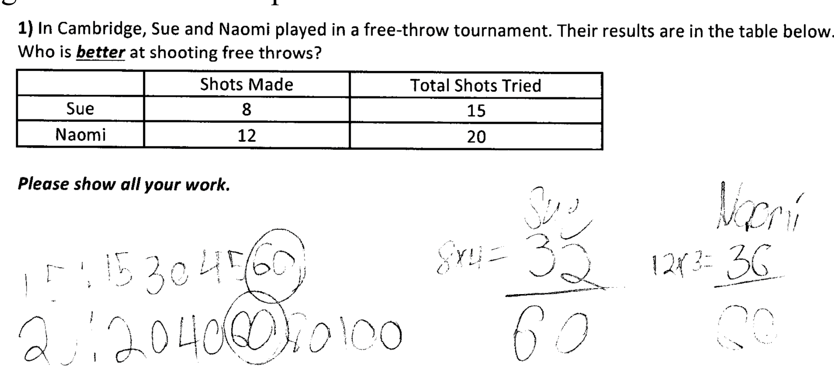

1) Shani and Keisha have both set up lemonade stands. Shani's lemonade recipe uses 2 cups of lemon juice and 1 cup of water. Keisha's lemonade recipe uses 3 cups of lemon juice and 2 cups of water.

Shani's Lemonade	Keisha's Lemonade

Whose lemonade tastes ***more*** "lemony"?

Show all your work.

2. **LCM**
   a. The MOST CORRECT answer for all problems (except for problems asking about check marks)
   b. Finding a least common denominator and making ratios with the same denominator
   Figure B2. LCM Example

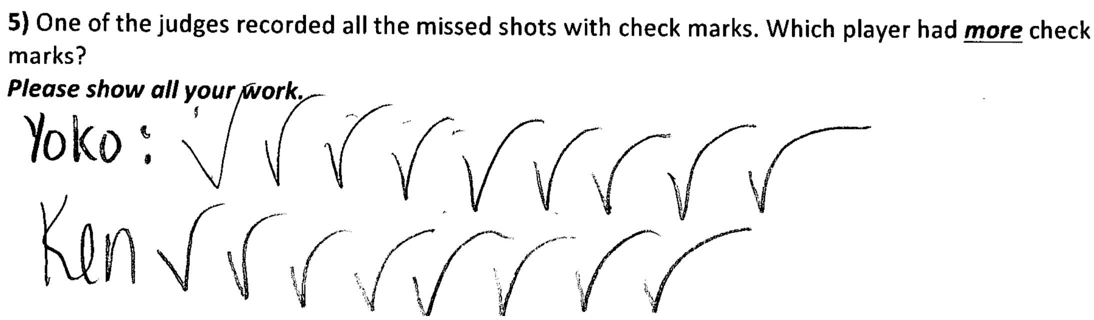

1) In Cambridge, Sue and Naomi played in a free-throw tournament. Their results are in the table below. Who is ***better*** at shooting free throws?

	Shots Made	Total Shots Tried
Sue	8	15
Naomi	12	20

Please show all your work.

3. **Subtraction**
    a. Normal arithmetic subtraction
    b. Note: Sometimes may draw check marks that do not look like subtraction but are
        i. Ex: Answer is 25-16=9 and 20-12=8, but the student instead drew nine check marks and 8 check marks. This would be coded as subtraction.

Figure B3. Subtraction Example



        ii. Note: Student may do addition but really be subtracting.
            1. Ex: may write 9+16=25. This would be coded as subtraction.
            2. For this reason, Always count what the student does/look closely at their work

4. **Ratio**
    a. Code ratio if student wrote out problem as the three different ways to write a ratio
        i. 3/5
        ii. 3:5
        iii. 3 to 5
    b. ONLY use ratio if the numbers are set up like this but no work is done with them(just because you see 3/5 doesn't mean it's a ratio, check to see if they tried to get the same common denominators—because that would be LCM)
        i.

5. **.Valid Strategy**
    a. A strategy is considered valid if the student's setup could give them the correct answer from their work.
        i. If the student thought about something that would work that we had not thought about
        ii. Pretty rare to code something as a 5

6. **Invalid Strategy**
    a. Anything the student did that is not one of the other codes and would not count as a 5
    b. Examples: addition, partially rewriting the problem, multiplication, drawing a diagram, using a less than or greater than sign, etc.
        i. Anything that may seem odd and have no correlation to the other codes
        ii. Generally will use this code a lot

7. **Rewriting the problem.**
    a. When the problem is rewritten EXACTLY as it is shown
    b. If leave certain parts out or only rewrite certain parts of the problem, it IS NOT rewriting the problem.
    c. For word problems, this would mean rewriting the words very close to verbatim
    d. For chart problems, this would mean redrawing the chart exactly as is shown
        i. Exception: If they redraw the chart as is shown but add anything to it (color in cups differently, shade only half the cups, write numbers on top of each other instead of next to each other as is shown in the original problem, etc.) then it is NOT considered rewriting the problem
        ii. Usually when the problem is only partially rewritten, it is coded as a 6, unless the student did some kind of arithmetic that can be coded as something else
        iii. Example- Here to student rewrote the majority of the problem with only adding the word "because" which did not help to explain how he was thinking in any way. This would be coded as a 7 because he repeats the problem extremely close to verbatim.

8. **Cross Multiplication**
    a. When the student sets up ratios and cross multiplies the diagonal numbers
        i. Sometimes will cross multiply without setting up the ratios (Student will just refer to the numbers in the table). This still counts as cross multiplication.
        ii. Note: ALWAYS check to see if the student attempted to do LCM or any type of division, as this would be more "correct" to code as than to code as cross multiplication

9. **999. Nothing is written**
    a. Either for the work or for the answer
    b. NOTHING IS WRITTEN (if in the work space they write "I don't know" this would still be coded as a 6 because there is *something* there).
    c. If student has nothing in the work space, code as 999, even if student wrote something on the contender line. However, if student wrote "I don't know" or drew a cross out or something, code as 6, not 999

Final Notes

- Always code for the MOST correct answer
    - Order of correctness: LCM, Division, Cross Multiplication, Ratio, Subtraction, Other Valid Strategy, Invalid Strategy, Rewriting the Problem, No answer

    - this is the order for all problems except "check mark" problems where the most correct answer is Subtraction
- If a student crosses out "correct" strategies and uses a different one to find their answer, still code for the MOST correct strategy they used, even if it is crossed out
  - If a student does LCM and crosses it out and then does subtraction and gets their answer from the subtraction, still code for LCM even though it is crossed out
- If student does one thing in his/her work, but in writing describes something else, code for whichever is most correct. Example, if student shows subtraction work, but then writes " but when I did the fractions Nia was best", code as ratio
- Always pay attention to the TYPE of problem you are coding
  - Different problems have different "correct" answers
    - Check mark problems- Most correct answer is subtraction
    - All other problems-Most correct answer is LCM

- Do not take off for spelling (eg if student misspells contender's name, still code as correct if it is clear who the student means)

- If student answers a question by pointing to something else, code as what they are pointing to. For instance, in the second part of the Ken/Yoko problem if student refers to work done in 1st part of problem, give credit
-

Ratio Code

Overview:

1 = mentions comparison or a relationship between two quantities or numbers, etc. (does not have to be the exact definition, but must relate two things to each other or express the idea of proportion). If a student gives an accurate or valid description of 'ratio' - even if it is not quite the definition we provided or is a more specific example (e.g., the rate of something; the odds of something happening) - code as 1.

0 = does not mention comparison or relationships at all (this includes things like, "The difference between two numbers", since "difference" is not proportional), or otherwise does not express the idea of proportion.

999 = no answer/blank

Details

- A correct definition of ratio must include **relationship** and a **correct number concept.** Correct number concepts include: number, amount, quantiy (accept any spelling or quality). Incorrect number concepts that would not be coded as 1 include: anything refering to addition or subtraction, pattern, something/other.

- If the student provides a correct example, code it as 1.
- If the student answers rate, fraction, or percent, code it as 1.
- If a student defines ratio in the context of the Ken and Yoko problem, code it as 1. [Ex/ "how many times you attempted something and how many times it worked.