Spring 5-2009

The Role of Gender Identity on the Effects of Stereotype Threat: An Examination of Girls' Math Performance in a Single-sex Classroom

Erin E. Twamley
Macalester College, etwamley@macalester.edu

Running Head: GENDER IDENTITY AND THE EFFECTS OF STEREOTYPE THREAT

The Role of Gender Identity on the Effects of Stereotype Threat:

An Examination of Girls' Math Performance in a Single-sex Classroom

Erin Twamley

Macalester College

May 2009

Abstract

Despite all of the advancements women have made in the field of mathematics, the negative stereotype regarding women's mathematical competence persists. Stereotype threat research demonstrates that the negative stereotype contributes to significant gender differences in attitudes, academic achievement, and educational and career attainment in math. The current longitudinal study focused on stereotype threat as an explanation for how a negative gender stereotype influences the mathematical performance of middle school girls in math in a single-sex setting. In particular, the study examines how the girls' gender identification moderates the effects of stereotype threat. The results of the study indicate that stereotype threat does operate in a single-sex setting and that participants' math performance varies significantly depending on their overall gender identification.

The Role of Gender Identity on the Effects of Stereotype Threat:

An Examination of Girls Math Performance in a Single-sex Middle School

As early as 1st grade, many girls report that boys are better at math than girls (Ambady, Shih, Kim, & Pittinsky, 2001; Furnham, Reeves & Budhani, 2002; Jacobs, 1991; Klebanov & Brooks-Gunn, 1992; Lummis & Stevenson, 1990; Shih, Pittinsky, & Ambady, 1999). Despite all of the advancements that women have made in the field of mathematics, the negative stereotype regarding women's mathematical abilities and competence persists. This negative stereotype is reinforced as research reveals that there are significant gender differences in academic achievement, and degree and professional attainment in math. However, these gender differences may not only reinforce the negative stereotype, but the negative stereotype itself contributes to the actual gender differences in math. Accordingly, psychology research has examined the onset, development, and persistence of gender differences in the field of mathematics.

A growing body of social psychological research focusing on a phenomenon known as "stereotype threat" provides the foundation for understanding the role of the negative stereotypes in the gender achievement gap in mathematics. Stereotype threat research demonstrates that the awareness of the negative stereotype *threatens* women and provokes women's underperformance compared to men (Freeman, 2003; Huguet & Regner, 2007; Johns, Schmader, & Martens, 2005; Keller, 2002; Onwuegbuzie, 1995; Shih et. al., 1999; Schmader, 2002; Schmader & Johns, 2003; Smith & White, 2002; Smith, Sansone & White, 2007; Spencer, Steele & Quinn, 1999; (Steele, 1998).[1] "Stereotype threat" describes a situational phenomenon in which individuals are *threatened* by their awareness of a negative stereotype and how this awareness interferes with their two primary achievement motivation goals in the situations (Beilock, McConnell & Rydell, 2007).[2] Individuals' achievement motivation goals in stereotype threat situations are to avoid fitting the negative stereotype and perpetuating the negative stereotype by performing incompetently (Osborne, 2001). The performance of the individual in stereotype threatening situations reflects not only the individual's competence, but the social group to which the individual belongs. Furthermore, individuals' motivation to perform well is essential in fulfilling the need for positive self evaluation and validation (Harkins, White, & Utman, 2000), and thus awareness of the stereotype proposes a *threat*. The present study investigates how awareness of a negative gender stereotype, the gender composition of the setting, and the social identification of girls influence girls' performance on a math test.

Academia and Single-sex Education

Single-sex education has a long controversial history within the US. Traditionally single-sex education has been used to serve privileged groups such as men or upper class women within the US. As a result, movements toward the mainstreaming of co-ed public education within the US began in the 1970's resulting in the significant decline of single-sex education (Wolf-Wendel, 2008). Most single-sex education models are found in parochial, private or higher educational institutions. However, a growing contemporary school of thought has emerged reconsidering and even advocating for the use of a public single-sex model as a legitimate means to educating students (Wolf-Wendel, 2008). Research since the 1980's, research has found numerous positive effects for girls and boys of color in single-sex education models, including positive self-image, diverse career choices, stable occupational achievement patterns, and positive attitudes toward academics (Hubbard & Datnow, 2005; Lee & Bryk, 1986; Mael, Alonso, Gibson, Rogers, & Smith, 2005; Sneed, 2007). In particular girls reported more positive attitudes, achievement, educational and career achievement in typically "male" domains such as math (Hubbard & Datnow, 2005; Mael, Alonso, Gibson, Rogers, & Smith, 2005; Sneed, 2007).

In addition, recent stereotype threat research has found that individuals' mathematical performance in stereotype threat situations is moderated by the gender of the setting; with women performing optimally in single-sex settings (Huguet & Regner, 2007; Inzlicht & Ben-Zeev, 2000; Keller & Molix, 2007; Marx & Stapel, 2006; Neuville & Croizet, 2007).[3] All of this research suggests that this setting offers unique qualities that are critical in understanding how stereotype threat operates and, as a result, the current study examines stereotype threat in a single-sex setting.

There are several facets in gender composition research including the latest research examining the salience of gender identity in single-sex versus co-ed settings. Previous work indicates that gender identity is formed around differences between the sexes, differences that are explicitly visible and differences prescribed in roles, attitudes and behaviors of the sexes (Shaw, 1995). Thus, the gender composition of a setting, (i.e. a single-sex), lacks both the visible and prescribed gender differences, as there is no social comparison group (i.e., boys) (Brutsaert, 1999; Keller & Molix, 2007; Shaw, 1995). Consistent with this proposal are the results of studies finding that the salience of gender identity in a single-sex setting is minimal (Brutsaert, 1999; Deaux & Major, 1987; Keller & Molix, 2007) and typically girls in single-sex settings do not identify with their gender (Brutsaert, 1999). This research demonstrates that the pervasiveness of gender identity is moderated by the gender composition of a setting.

Gender Identification and Stereotype Threat

The literature proposes a critical relationship between the gender composition of a setting and gender identity while supplementing emerging stereotype threat research by incorporating an understanding of individuals' gender identity and the effects of stereotype threat. The initial work of Steele (1998) theorized that an individual's identification with their gender (social self) links them to the negative stereotype and, thus, *threatens* their performance. The literature of Marx and Stapel (2006) which compiles numerous studies including: Shih et. al., (1999) and Schmader (2002), provides support and evidence for the role of social identity in stereotype threat situations.

This research is grounded in an understanding of Tajfel's (1981) Social Identity Theory. Social Identity Theory posits there are two types of identities, personal and social; personal identities are derived from unique characteristics, behaviors and beliefs that comprise the individual, whereas social identities are formed through the identification with or membership in specific categorical social groups (i.e. gender) (Tajfel, 1981; Tajfel & Turner, 1986). Individuals strive to maintain both positive personal and social identities.

Social Identity Theory was first applied to stereotype threat by Steele (1998): individuals identification with their social self (i.e. gender) links them to the negative stereotype, *threatening* their performance. Shih et al., (1999) study unintentionally found support for Steele's (1998) proposal. Shih et al., (1999) examined participants who had two conflicting stereotypes (one positive and one negative) regarding their mathematical performance, i.e. Asian women. There is a positive stereotype regarding Asian's mathematical competence indicating that they should perform optimally while on the contrary, there is a negative gender stereotype, that because they are women they should underperform. The study found that depending on the identification of participant, as either Asian or as a woman, her math performance significantly varied. The women who identified as Asian optimally performed, while the women who identified as women underperformed. Schmader (2002) intentionally examined how strongly gender identified women performed on mathematical tasks compared to weakly gender identified women finding that women who identified strongly with their gender underperformed significantly. Finally, Marx and Stapel (2006) synthesized the application of Social Identity Theory to stereotype threat situations contending that gender identification is critical to understanding the effects of stereotype threat. Individuals' gender identification in stereotype threat situations is critical in two aspects: the gender identity is linked to the negative stereotype threatening the individual, and the negative stereotype conflicts with the individuals' attempts to maintain positive social identities.

Numerous subsequent studies have found that individuals are linked to the negative stereotype through their identification with the social identity that the negative stereotype regards (i.e. gender) in stereotype threat situations (Good et. al., 2003; Marx & Stapel, 2006; Marx et. al., 1999; Osborne, 2001; Schmader 2002; Steele, 1998; Shih et. al., 1999; Wheeler & Petty, 2001). Typically studies extending this literature have conditioned participants to (dis)identify with a social identity that is linked to the negative stereotype (Marx et. al., 2005; Marx & Stapel, 2006; Neuville & Croizet, 2007; Schmader, 2002; Wilson & Liu, 2003). The findings of these studies support the theory; women in stereotype threat conditions, who identify as a woman, underperform compared to women who do not identify as a woman (Ben-Zeev & Inzlicht, 2000; Marx & Stapel, 2006; Neuville & Croizet, 2007; Schmader, 2002; Wout, Danso, Jackson & Spencer, 2008).

Based on these studies, researchers have been able to make a vital claim in stereotype threat work: the social or personal identification of individuals in stereotype threat situations moderates performance. Moreover, the identification of these individuals can be situational, activated by the gender composition of a setting. To incorporate these findings of current research, the present study measures individuals' gender identification in a single-sex stereotype threat situation to examine the effects of stereotype threat.

The Current Study

The negative stereotype regarding women's mathematical competence is prevalent within U.S. culture (Maccoby & Jacklin, 1974), media (Eccles & Jacobs, 1991), parental beliefs (Frome & Eccles, 1998; Furnham, Reeves, & Budhani, 2002; Herbert & Stipek, 2005; Hyde et. al., 2008; Li, 1999; Smith & Hung, 2008) and teacher's beliefs (Eccles-Parson, 1984; Fennema et. al., 1990; Li, 1999). Teachers' beliefs have been exposed revealing that typically they view math as a male domain and as a result, teachers' over estimate male's ability to do math, maintain higher expectations for male students (Eccles-Parson, 1984; Fennema et. al., 1990; Li, 1999), and interact with male students in the classroom at a higher rate (Jungwirth, 1991; Fennema & Sherman, 1977). Both parental and teachers' beliefs are critical as such beliefs are incorporated into the academic feedback children begin to receive and rely on in middle school. This academic feedback shapes students' beliefs about their academic abilities, strengths, weaknesses and academic achievement (Eccles & Wigfield, 1994; Frome & Eccles, 1998 Herbert & Stipek, 2005). Moreover, studies reveal that the stereotypical beliefs of mothers are correlated to their daughter's competence beliefs and actual academic achievement (Bleeker & Jacobs, 2004; Eccles, 1993; Frome & Eccles, 1998; Jacobs & Eccles, 1992; Klebanov & Brooks-Gunn, 1992; Wigfield & Eccles, 1994; Wigfield et. al., 1997). Consequently, the negative gender stereotype can be integrated into students' academic feedback and subsequently, their competence beliefs and math performance.

The findings of stereotype threat research affirm that the negative stereotype continues to contribute to women's underperformance in math (Huguet & Regner, 2007; Schmader & Johns, 2003; Spencer, Steele & Quinn, 1999). Emerging evidence in the growing work of stereotype threat and women's mathematical performance indicates that there are two critical components that need to be considered: 1) gender composition of a setting (single-sex v. co-ed) shown to moderate the effects of stereotype threat (Huguet & Regner, 2007; Keller & Molix, 2007; Marx & Stapel, 2006) and 2) participants' social gender (dis)identification shown to moderate the effects of stereotype threat (Ben-Zeev & Inzlicht, 2000; Marx & Stapel, 2006; Neuville & Croizet, 2007; Schmader, 2002; Wout, Danso, Jackson & Spencer, 2008). Moreover, there is a critical relationship between the gender composition of a setting and individuals' gender identification in the setting (Brutsaert, 1999; Keller & Molix, 2007; Shaw, 1995).

To address this emerging research and highlight the important contribution of stereotype threat to our understanding of the gender achievement gap in math, the current study was conducted. The study examines middle school (5^{th} and 6^{th} grade) girls' gender (dis)identification in a single-sex classroom and the effects of this gender (dis)identification based on induced stereotype threat.

Based on the work of Inzlicht & Ben-Zeev (2000), the first aim of this study was to measure gender identification of middle school girls in a single-sex setting (Liu & Wilson, 2003). Drawing on previous research indicating that the gender composition of a classroom may influence the salience of gender identity (Brutsaert, 1999), girls' gender identification in the single-sex setting was expected to be low.

The second purpose was to examine how the middle school girls' (dis)identification with their gender in a single-sex setting affects stereotype threat by examining the effects on the girls' mathematical performance and anxiety levels. Previous work (Marx & Stapel, 2006; Schmader, 2002) indicates that if the girls have low gender identification, girls should perform competently on their math tests despite the induction of stereotype threat.

To test these predictions, a longitudinal study was conducted using middle school girls in a single-sex math classroom. The participants completed the study three times throughout the 2008 - 2009 academic year.[4]

Method

Participants

All of the middle school girl participants (*N* = 73) in this study had parental/guardian consent.[5] Over 70% of the families contacted agreed to have their child participate in the study. Participants ranged in age from 10 to 12 years old, were in 5th grade (*N* = 45) and 6th grade (*N* = 28), and 54 % of participants' identified themselves as a person of color while 46% as White/Caucasian (See Appendix A).[6] Seventy percent of the participants reported that they liked math while 12% reported that they did not like math. All of the participants attended Laura Jeffrey Academy a girls focused middle school in St. Paul, MN. Most of the participants reported previously attending a co-ed school; only one participant reported being home schooled, and no participant reported attending a single-sex school.

Procedure

Modeled after the most common model used to replicate or induce stereotype threat (Steele, Aronson & Spencer, 2002) the current study asked middle school girls to perform on a math test.[7] Participants were assessed at three points in time at eight week intervals. The study occurred in a single-sex math classroom at Laura Jeffrey Academy. All the participants were randomly pre-assigned a code number coordinated with the study materials at each session. A master list containing the participants' names and code numbers was maintained by the female researcher to ensure that individuals' participation in the study remained confidential.

There were non-participants, participants, the two classroom teachers and the female researcher present during the study. The female researcher instructed, administered and collected the materials in this study. All of the students were instructed by the female researcher that there were numerous math "activities" to be completed.[8] The non-participants completed a task assigned by the teacher while the participants received and completed the packet of materials for this study. The materials for the study were distributed in 1 of 6 colors to prevent the participants and non-participants from knowing who was doing what mathematical "activity."[9]

Participants were randomly assigned to a (1) stereotype threat condition or (2) non stereotype threat condition. Participants in the *stereotype threat* condition ($N = 38$) were asked to circle their gender prior to completing the math test. This is a common technique used to induce stereotype threat (Marx & Goff, 2005; Spencer, Steele & Quinn, 1999; Steele, Spencer & Aronson, 1995).[10] Participants in the *non stereotype threat* condition ($N = 36$) were not prompted to circle their gender prior to completing the math test.

The longitudinal study occurred at three sessions: session A in October 2008, session B in December 2008 and session C in February 2009. All of the participants completed the Learn about You Survey at the initial session (A) to provide demographic information (See Appendix B). All of the participants completed the study materials in the following order, the 15 problem math test, the Questions for You Survey and the Thoughts Survey at all three sessions (A, B, C). All three of the math tests consisted of 15 difficult and easy mathematical problems appropriate for 5th and 6th graders from the Texas Assessment of Knowledge and Skills (2003, 2006) (See Appendix C). The Questions for You survey consisted of 2 scales, the State Anxiety Inventory for Children (STAIC-S) (Spielberger et. al., 1973) and the Explicit Stereotype Threat Scale

(Marx & Goff, 2005) (See Appendix D). The Thoughts Survey consisted of a 6 item gender identification scale modified from Wilson & Liu's (2003) study (See Appendix E).

At the final session (C), participants and non-participants were debriefed in the classroom. The purpose of the study, the hypotheses and an explanation of the information collected was discussed. Participants were also provided with a supplemental sheet to discuss with parents if needed. Participants were encouraged to ask questions and discuss the study at this time.

Measures

Texas Assessment of Knowledge and Skills. The Texas Assessment of Knowledge and Skills (TAKS) is a standardized test used in the state of Texas to assess the mathematical skills and progress of elementary school students. The test was first used in the 1990-1991 academic school year and has since been adapted and used every year. The test is commonly used for 3rd through 6th graders with a variety of skill level math problems that vary in difficulty. Fifteen math problems were randomly selected for each math test from the original 90 math problems from TAKS (2003, 2006). Each math test at each session contained both difficult and easy problems involving geometry, division, metric system, fractions and multiplication specifically for 5th and 6th graders. The format of this mathematical test is consistent with other math tests used in stereotype threat studies to assess math performance (Pezdek, Berry & Renno, 2002). The use of both easy and difficult problems helps to account for the variability of math skills in the participants, but also individuals in stereotype threat conditions underperform on difficult problems compared to easy ones (Huguet & Regner, 2007). The math performance of participants was compiled using accuracy scores.[11] Each participant received an accuracy score:

computed by taking the number of math problems correct divided by the number of the math problems completed, for each of the three math tests completed.

The State-Trait Anxiety Inventory for Children (STAIC). Participants were asked to complete the STAIC-S in the Questions for You Survey after completing the math test. The State-Trait Anxiety Inventory for Children was developed by Spielberger et.al (1973) to measure two types of anxiety, Trait Anxiety and State Anxiety in children.[12] The 40 item scale uses a 3 point rating scale. In this particular study the STAIC S-Anxiety items, a subsection of the STAIC, were used. The STAIC S-Anxiety is a twenty item measure of how children feel at that moment. In the current study, it was used to measure how anxious the participants felt after completing the math test. Numerous studies have used the scale to measure anxiety in children (Hedl, & Papay, 1982; Houston, Fox & Forbes, 1984; Papy, Costello & Hedl, 1975; Psychountaki et.al., 2003). Participants were asked to complete the statement "I feel…" with a choice of 1 of 3 words, examples include very upset, upset, not upset. Responses to the STAIC-S were coded using a 3 point scale. Answers referring to key terms such as upset, worried or nervous were valued 3, 2, 1 whereas answers that reflected key terms such as calm, pleasant and relaxed were scored 1, 2, or 3. Each participant had a raw anxiety score between 20 and 60 at each session. The Cronbach's alpha for the scale at each sessions were: session A, is $\alpha = .90$, session B, $\alpha = .89$, and session C, $\alpha = .92$.

Explicit Stereotype Threat Scale. The ESTS scale was completed by participants after the math test. The Explicit Stereotype Threat Scale (ESTS) was first developed by Marx and Goff (2005) as a manipulation check for their experiment examining participants' experience of stereotype threat. The scale has been replicated in numerous studies (Marx & Goff, 2005; Marx & Stapel, 2006; Davies, Goff & Steele, 2008). In the current study, the ESTS was used to measure the experience of stereotype threat in the single-sex setting. The ESTS is a four item scale that has been replicated to fit several experiments. Example items include "I worry that my ability to perform well on standardized tests is affected by my being a girl" to which participants respond on a scale ranging from 1 (strongly disagree) to 7 (strongly agree). Responses to the ESTS were averaged with a high average indicating a strong experience of stereotype threat. The Cronbach's alpha for the scale in each session was: session A, is α =.82, session B, α = .58, and session C, α =.84.

Gender Identification Scale. The Gender Group Identification Scale used in this study was adapted from Wilson & Liu (2003). This scale was used in the Thoughts for You Survey that participants completed at the end of the study. The Gender Group Identification Scale has been used in other studies to assess individual's identification with their gender identity. This six item scale was modified from the original items to be context specific, for example an item such as "I often think of myself as a girl" was formatted to "I often think of myself as a girl in school." Participants were asked to respond to such items on a scale from 1 (strongly disagree) to 5 (strongly agree). The Cronbach's alpha the scale in each session was: session A, is α =.60, session B, α = .65, and session C, α = .73.

The gender identification of participants varied within each session (A, B and C) as well across time and, thus, the gender identification of participants was both a between and within subjects variable. To address this issue the gender identity scores of each participant at each Session (A, B, C) were combined into one overall gender identity score for each participant. The overall gender identity score was calculated in 3 steps. The first step was to compute a raw gender identity score by averaging participants' responses to the 6 item scale. Each participant had three raw gender identity scores (one for each session) that were of a numerical value between 0 and 6. Each of these raw gender identity scores were converted to a value within a range: 0 – 3.99 as weak gender identification or 4- 6 indicating strong gender identification. In the final step, the gender identification of each participant, strong or weak at each Session (A, B, C), were summed to give a final gender identification. Participants' overall gender identification either remained strong, weak or changed throughout the study.

Results

Each participant had three dependent variables, namely accuracy, anxiety, and ESTS scores at session A, B and C; thus, each participant had in total nine scores that were compiled. Participants also had an overall gender identification that was computed and determined as one of three levels, strong gender identity, weak gender identity or changed gender identity.

Cross-sectional Analyses

Table 1.

Descriptive Statistics for Dependent Variables at Session A (October 2008)

	Condition			
	Stereotype Threat		Non-Stereotype Threat	
Dependent Variable	*M*	*SD*	*M*	*SD*
Accuracy	0.68	0.21	0.73	0.18
Anxiety	1.09	0.30	1.03	0.18
ESTS	2.00	1.07	2.40	1.43

Note. Lower means for accuracy scores indicates lower math performance. Anxiety means closer to 1 indicate low anxiety and closer to 2 indicate high anxiety. Higher means for the ESTS indicates participants experienced stereotype threat.

Table 1 shows the means and standard deviations for each of the three variables at session A (October 2008). Accuracy scores are interpreted based on percentage correct, with participants in the stereotype threat condition correctly completing 68% of the math problems while participants in the non-stereotype threat condition correctly completing 73% of the math problems. Although the difference was not significantly significant F (1, 66) = .036, p > .05, the accuracy scores were in the expected direction based on previous research finding that participants in stereotype threat conditions underperform compared to participants in non-stereotype threat conditions (Huguet & Regner, 2007; Onwuegbuzie, 1995; Schmader & Johns, 2003; Spencer, Steele & Quinn, 1999; Steele, 1998). Participants' in stereotype threat condition reported relatively higher anxiety score (1.09) than participants in non-stereotype threat condition (1.03).[13] Although, the difference in participants' anxiety based on condition was not statistically significant F (1, 61) = .985, $p > .05$, the anxiety scores were in the expected direction based on previous research showing that participants in stereotype threat conditions generally report higher anxiety (Keller & Dauenheimer, 2003; O'Brien & Crandall, 2003; Onwuegbuzie, 1995; Pekrun et. al., 2007; Spencer, Steele & Quinn, 1999; Steele & Aronson, 1995). Although, participants in the non-stereotype threat condition reported experiencing stereotype threat more than participants in stereotype threat condition, the difference was not significantly significant F (1, 65) = 1.69, $p > .05$.

Table 2.

Descriptive Statistics for Dependent Variables at Session B (December 2008)

	Condition			
	Stereotype Threat		Non-Stereotype Threat	
Dependent Variable	*M*	*SD*	*M*	*SD*
Accuracy	0.62	0.20	0.79	0.23
Anxiety	1.15	0.36	1.06	0.25
ESTS	1.91	0.99	2.70	1.08

Note. Lower means for accuracy scores indicates lower math performance. Anxiety means closer to 1 indicate low anxiety and closer to 2 indicate high anxiety. Higher means for the ESTS indicates participants experienced stereotype threat.

Table 2 shows the means and standard deviations for each of the three variables at session B (December 2008). The accuracy scores of participants' were again in the expected direction with participants in stereotype threat condition correctly completing fewer problems

(62%) then participants in the non-stereotype threat condition (79%), although the difference was not significantly significant F (1, 61) = 1.45, $p > .05$. The anxiety scores were in the expected direction with participants in stereotype threat condition reporting higher anxiety (1.15) than participants in the non-stereotype threat condition (1.06), although the difference was not statistically significant F (1, 61) = 1.13, $p > .05$. [13] Participants in non-stereotype threat condition reported experiencing stereotype threat more than participants in stereotype threat condition and the difference was not statistically significant F (1, 61) = 9.71, $p > .05$.

Table 3.

Descriptive Statistics for Dependent Variables at Session C (February 2009)

	Condition			
	Stereotype Threat		Non-Stereotype Threat	
Dependent Variable	*M*	*SD*	*M*	*SD*
Accuracy	0.61	0.27	0.58	0.25
Anxiety	1.91	0.39	1.10	0.31
ESTS	1.98	1.20	2.40	1.56

Note. Lower means for accuracy scores indicates lower math performance. Anxiety means closer to 1 indicate low anxiety and closer to 2 indicate high anxiety. Higher means for the ESTS indicates participants experienced stereotype threat.

Table 3 shows the means and standard deviations for each of the three variables at session C (February 2009). Participants in stereotype threat condition correctly completed more math problems (61%) than participants in the non-stereotype threat condition (58%), although the difference was not significantly significant F (1, 61) = .819, $p > .05$. Participants in stereotype threat condition had an average higher anxiety score (1.91) than participants in the non-stereotype threat condition (1.10) although, the difference in participants' anxiety based on condition was not statistically significant F (1, 61) = .843, $p > .05$. [13] Participants in non-stereotype threat condition reported experiencing stereotype threat more than participants in stereotype threat condition, although the difference was not significantly significant F (1, 61) = 1.65, $p > .05$.

In examining the descriptive statistics of the data at all three sessions, data for Session A and B are in the expected direction, but not at Session C.[14] This may be the result of a confound because students became aware that they were completing their last math test at Session C from their teachers prior to the study. Thus, the students may not have been motivated to perform well, but to simply finish the materials to learn what the study was about. As a result of this confound, only participants' data from Session A and B were analyzed, resulting in 65 of the 73 participants who completed the study providing data for subsequent longitudinal analyses. Therefore, the gender identity scores of participants were again recalculated. The gender identification of each participant, strong or weak at only Session A and B were combined to give new overall gender identification for only these two sessions. Participants' overall gender identification either remained strong, weak or changed from Session A to B. This new overall gender identification score for each of the participants was used in the longitudinal analysis of the results.

Longitudinal Analyses

To analyze the data at Session A and B, three mixed model analyses of variance were conducted to examine how the dependent variables varied across time (8 weeks). The between-subjects factors were condition, with two levels of stereotype threat or non-stereotype threat, and overall gender identification of the participants across time (strong, weak, changed).

Repeated measures analysis of variance (RMANOVA) for accuracy scores at Session A and B revealed that there were no main effects for accuracy scores, F (1, 52) = .298, $p > .05$; and condition, F (1,52) = 2.71, $p > .05$. The RMANOVA did reveal a significant main effect for the overall gender identification, *F* (2, 52) = 3.40 $p < .05$. Participants' performance varied significantly depending on their overall gender identification (strong, weak or changed). To determine which overall gender identification groups' performance varied significantly from one another, a Tukey HSD post hoc was conducted. The results of the Tukey HSD post hoc found a significance difference in the math performance between strongly and weakly gender identified participants, changed and weakly gender identified participants and changed and strongly gender identified participants, p =.05. Participants that identified strongly (*M* =.648) performed the worst compared to both the weakly identified participants (*M* = .792) and participants who changed gender identification (*M* = .72) (See Figure 1).

The RMANOVA revealed no interaction effects between accuracy scores and overall gender identification, *F* (2, 52) = .373, $p > .05$, and accuracy scores, condition and overall gender identification, *F* (2, 52) = .103, $p > .05$. The RMANOVA did reveal a significant interaction effect between participants' accuracy scores and condition *F* (1, 52) = 9.10, $p < .05$. The accuracy scores of participants based on their condition were significantly different at Session A and Session B (See Figure 2).

The RMANOVA for anxiety of participants' at Session A and B revealed no main effects for anxiety, *F* (1, 52) = .901, $p > .05$, condition, *F* (1, 50) = 1.08, $p > .05$, and overall gender identification *F* (1, 52) = .438, $p > .05$. The RMANOVA revealed no interaction effects between anxiety and condition, *F* (1, 52) = .141, $p > .05$, anxiety and overall gender identification, *F* (2, 52) = .743, $p < .05$, or anxiety, condition and overall gender identification, *F* (2, 52) = 2.01, $p > .05$.

The results of the RMANOVA for experience of stereotype threat at Session A and B revealed no main effects experience of stereotype threat *F* (1, 51) = .015, $p > .05$, and overall gender identification, $F(2, 51) = 1.13$, $p > .05$. The RMANOVA did reveal a significant main effect for condition, $F(1, 51) = 6.88$, $p < .05$. Participants in non-stereotype threat condition reported experiencing stereotype threat significantly more than participants in stereotype threat condition. The RMANOVA revealed no significant interaction effects between experience of stereotype threat and condition, $F(1, 51) = 1.15$, $p > .05$, experience of stereotype threat and overall gender identification, $F(2, 51) = .045$, $p > .05$, or experience of stereotype threat, condition and overall gender identification, $F(2, 51) = .218$, $p > .05$.

Discussion

Stereotype threat (Steele, 1998) has been used to explain women's underperformance in math (Huguet & Regner, 2007; Pezdek, Berry & Renno, 2002; Schmader & Johns, 2003). The current study was an attempt to extend stereotype threat research by examining two key components: 1) if the gender composition of a single-sex setting moderates the effects of stereotype threat; and 2) if individuals' gender (dis)identification moderates the effects of stereotype threat. Therefore, the quasi-experiment was conducted to examine the effects of stereotype threat on middle school girls' math performance and measure the girls' gender identification and its impact (if any) on the girls' math performance.

Both the anxiety scale (Spielberger, 1973) and the Explicit Stereotype Threat Scale (Marx & Goff, 2005) were used in the current study as potential indicators of stereotype threat. Prior stereotype threat research has shown that participants generally have high anxiety (Keller & Dauenheimer, 2003; O'Brien & Crandall, 2003; Onwuegbuzie, 1995; Pekrun et. al., 2007; Spencer, Steele & Quinn, 1999; Steele & Aronson, 1995) and report experiencing stereotype threat (Marx & Goff, 2005). In this study, the anxiety scores and the ESTS scores of participants did not reveal that stereotype threat was operating.

There were no statistically significant results for the RMANOVA conducted to examine the anxiety scores of participants. Anxiety measures are commonly used in stereotype threat studies as a potential indicator of stereotype threat as previous studies have found that participants in stereotype threat situations typically report having high anxiety.[12] For the current sample, the anxiety measure did not indicate higher anxiety for participants in stereotype threat condition. This may be a result of the artifact of the anxiety measure because of the participant ages and attention. Although, the anxiety scale (STAIC-S) has proven to be a reliable and valid measure for children's anxiety, this is a 20 item monotonous measure. Participants, ten to twelve year old girls, were asked to respond repeatedly to finish the same statement, "I feel…" right after completing the fifteen problem math test. Participants may have been minimally motivated to complete the anxiety measure thoroughly.

RMANOVA results for the ESTS scale revealed only one significant main effect for condition for the experience of stereotype threat measure was found. Participants in non-stereotype threat condition reported experiencing stereotype threat significantly more than participants in stereotype threat condition. This is not consistent with previous research that uses the ESTS finding that individuals in stereotype threat conditions usually have higher ESTS scores (Marx & Goff, 2005; Marx & Stapel, 2006; Davies, Goff & Steele, 2008) suggesting that the validity and the reliability of the ESTS in the current study are questionable. The ESTS was used as a potential indicator of stereotype threat as previous studies have found that participants in stereotype threat situations typically can indicate they are experiencing stereotype threat. The ESTS in this study was a 4 item scale modified from the original 6 item scale typically used s a manipulation check for studies using college students (Marx & Goff, 2005). For the current study, the language and the items were modified in attempt to appropriately engage middle school participants, but the low reliability of these items indicates that further modification of the items is required. For the current sample, the ESTS was not a valid and reliable measure for the population.

Although, the ESTS and anxiety measure did not reveal that stereotype threat At the crux of this stereotype threat study is the examination of participants' math performance. The math performance of participants was measured and compared using accuracy scores. Accuracy scores are the percentage of correctly completed math problems. Each participant had a calculated accuracy score for each session (A and B). Based on previous stereotype threat research (Marx & Stapel, 2006) there were two main hypotheses: 1) the math performance of participants in stereotype threat condition would be significantly lower compared to those in non-stereotype threat condition, and 2) the math performance of participants who identified strongly with their gender would be significantly lower compared to participants who had identified weakly.

To examine the first hypothesis (girls in stereotype threat condition should underperform compared to girls in non-stereotype threat condition) the mathematical performance of participants based on condition and across time was tested using an RMANOVA. This test showed that there was an interaction effect between participants' accuracy scores and condition. In other words, the performance of participants varied significantly across time and across condition with participants' accuracy scores in stereotype threat condition significantly lower than participants' accuracy scores in non-stereotype threat condition. Participants in stereotype threat condition underperformed on the math test compared to participants in non-stereotype threat condition, consistent with previous stereotype threat research (Cohen & Walton, 2003). The effects of stereotype threat were evident in the significant difference between participants' mathematical performance based on condition indicating that stereotype threat was operating in the single-sex classroom.

Finally, to examine if the second hypothesis (girls with strong gender identification would underperform) was supported another RMANOVA was conducted to examine how the gender identification of participants (if at all) impacted their mathematical performance. The RMANOVA revealed no statistically significant interaction effects between accuracy scores and overall gender identification or accuracy scores, condition, and overall gender identification. Although, the RMANOVA results did reveal an overall main effect for participants' performance based on their gender identification, indicating that the math performance of participants varied significantly depending on their overall gender identification. To determine which gender identification groups (strong, weak and changed) performances' were significantly different a Tukey HSD post hoc test was completed. All of the gender identified groups, strong, weak and changed, varied significantly from one another. The math performance of participants was measured using accuracy scores. Participants' who had a strong gender identification performed significantly worse (64.8%) compared to participants who had weak gender identification (79%) or changed gender identification (72%). This finding is consistent with previous work (Good et. al., 2003; Marx & Stapel, 2006; Schmader, 2002; Wheeler & Petty, 2001) that contends individuals' gender identification (links them to the negative stereotype heightening the *threat* of the stereotype and thus, they underperform.

Although, the results were significant, it is important to note that this study, unlike previous studies, "measured," not conditioned individuals' gender identification. Therefore, participants' gender identification could be "changed" as the gender identification was measured at both sessions (A and B). The gender identification of participants varied at each session (A and B) and across sessions (A and B). Also indicating that the gender identity is cued by situational settings such as the gender composition of the classroom as previous work indicates (Brutsaert, 1999; Shaw, 1995). However, the variance within participants' and across time posed a significant problem in subsequent analyses, as the gender identification of participants in this study became both a between subjects and within subjects factor. As a result, the gender identity scores of participants had to be combined to make gender identity a between subjects variable. The results of this collapsing indicated that participants gender identification did not remain consistent throughout the study, rather gender identification of participants did change and not linearly.

The results of this study should be qualified by two major limitations: participant attrition rates and the lack of a comparative cohort. Participation in longitudinal studies is constrained due to basic attrition, especially when intended participants are students. Students often miss a class or even a day of school due to illness, vacation, detention and/or other circumstances. Consequently, collecting data from participants at three points in the study proved to be difficult. Although, initially there were 73 participants who completed the study at Session A, only 51 participants completed the study at all three sessions (A, B, C) and 65 for sessions A and B. Many participants failed to show up for a Session and/or did not respond to scales within the study and therefore, numerous data was thrown out. To increase participation more students should be recruited initially for the study and encouraged to be in the school on the day the study is to be completed. A larger population of participants at session A eliminates the seemingly significant effects of attrition and aid in the testing and analysis of the original proposed longitudinal (three sessions) experiment. Furthermore, the non-linear changes in participants overall gender identification indicate that to fully understand how gender identity is cued by situations and to compare the gender identification of individuals in various settings (i.e. single-sex versus co-ed) an analysis of data at three points in time is important.

In conclusion, the results of this study indicate that a comparative analysis of middle school students in single-sex versus co-ed setting is necessary to develop a thorough understanding of how gender identification and stereotype threat operates in the classroom. Research on gender identification and the effects of stereotype threat indicates that this is a critical component to be considered when evaluating the performance of individuals (Marx & Stapel, 2006). The findings of this study confirm that examining participants' gender identification is important in explaining the math performance of individuals in stereotype threat situations. Therefore, by extending this study to include a group of middle school girls within a co-ed setting (comparative cohort), one can examine more thoroughly how the gender composition of a setting impacts the gender identification of individuals. Such an inclusive study would provide more participants and a more complex understanding of the gender identification of individuals in co-ed and single-sex settings while examining the interaction of this identification on the effects of stereotype threat. This would provide a thorough analysis of how stereotype threat operates within the single-sex and co-ed setting while also allowing for the comparison of essential data.

The results of the present study support stereotype threat research while also extending stereotype threat research. This study indicates that stereotype threat is a phenomenon that exists outside of the laboratory setting and in real classroom settings. This is an important contribution as one of the major critiques and challenges of stereotype threat research is that most is research is conducted within a laboratory setting, and thus, little is known about how stereotype threat operates in real testing situations (Cullen, Waters, & Sackett, 2006). Moreover, gender identification of participants moderates their mathematical performance, as strong gender identification links the participant to the negative stereotype and these participants underperform. The results were in the expected direction and suggest that future research is necessary to produce statistically significant results.

The gender negative stereotype concerning women's mathematical competence is prevalent within U.S. culture (Maccoby & Jacklin, 1974), the media (Eccles & Jacobs, 1991), and within academia. As a result, a body of social psychological research, stereotype threat, examines how this negative stereotype contributes to the significant gender differences in math, as research indicates that there is a decline in girls' aptitude, attitudes and educational and career attainment in math beginning in middle school. These gender differences are alarming given the given the demographics of the population; women make up over 50% of US population (Smith & Hung, 2008) and nearly 50% of the students enrolled in higher education (Malcolm, 1986), but continue to be underrepresented in the field of math.

Further investigation into stereotype threat is crucial in broadening our understanding of the existing achievement gap between the genders in math (Muzatti & Agnoli, 2007). By focusing on the methodologies used to induce and examine the effects of stereotype threat we build a more comprehensive understanding of stereotype threat while expanding our conceptions of intelligence and academic performance by integrating and examining how situational factors such as social identity and negative stereotypes impact academic performance (Croizet & Quimzade, 2007; Lummis & Stevenson, 1990). Consequently, research examining why and how the gender achievement gap in math persists is crucial in order to eliminate this phenomenon.

References

Ambady, N., Shih, M., Kim, A., & Pittinsky, L. T. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. Psychological Science, 12(5), 385-390.

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

Blanton, H., Christie, C., & Dye, M. (2002). Social identity versus reference frame comparisons: The moderating role of stereotype endorsement. *Journal of Experimental Social Psychology, 38(3)*, 253 - 267.

Bleeker, M. M., & Jacobs, E. J. (2004). Achievement in math and science: Do mothers' beliefs matter 12 years later? *Journal of Educational Psychology, 96(1), 97-109.*

Brown, P., R., & Josephs, A. R. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology, 76(2),* 246-257.

Carpenter, D. M., Ramierez, A., & Severn, L. (2006). Gap or gaps: Challenging the singular definition of the achievement gap. *Education and Urban Society, 3(1),* 113-127.

Chipman, F. S., Krantz, D. H., & Silver, R. (1992). Mathematics anxiety and science careers among able college women. *American Psychological Society, 3(5),* 292-295.

Cohen, G.L., & Walton, G.M. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39(5),* 456-467.

Croizet, J., & Quiamzade, A. (2007). The social detriments of intellectual performance under threat: From interpersonal comparisons to the stereotype threat. *Swiss*

Journal of Psychology, 66(3), 139-144.

Cullen, J. M., Waters, D. S., & Sackett, R. P. (2006). Testing stereotype threat theory predictions for math-identified and non-math-identified students by gender. *Human Performance, 19(4),* 421-440.

Eccles, S. J., & Wigfield, A. (2002). Motivational beliefs, values and goals. *Annual Review of Psychology, 53,* 109-132.

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

Frome, M. P., & Eccles, S. J. (1998). Parents' influence on children's achievement-related perceptions. *Journal of Personality and Social Psychology, 74(2),* 435-452.

Furnham, A., Reeves, E., & Budhani, S. (2002). Parents think their sons are brighter than their daughters: Sex differences in parental self-estimations and estimations of their children's multiple intelligences. *The Journal of Genetic Psychology, 163(1),* 24-39.

Goff, A., P., Steele, M. C., & Davies, G. P. (2008). The space between us: Stereotype threat and distance n interracial contexts. *Journal of Personality and Social Psychology, 94(1),* 91-107.

Haplern, F. D. (1992). *Sex differences in cognitive abilities* (2nd ed.). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Harkins, G. S. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of Personality and Social Psychology, 91(3),* 436-455.

Harkins, G. S., & Jamieson, P. J. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology, 93(4),* 544–564.

Harkins, G. S., White, H. P., & Utman, H. C. (2000). The role of internal and external sources of evaluation in motivating task performance. *Personality and Social Psychology Bulletin, 26(1),* 100-117.

Herbert & Stipek. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Applied Developmental Psychology, 26(1),* 276-295.

Hubbard, L., & Datnow, A. (2005). Do single-sex schools improve the education of low-income and minority students? An investigation of California's public single-gender academies. *Anthropology and Education Quarterly, 36(2),* 115-131.

Huguet, P., & Regner, I. (2007). Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology, 99(3),* 545-560.

Hyde, J.S., Fennema, E., Lamon, S., (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin 107 (2),* 139-155.

Hyde, S. J., Lindberg, M. S., Linn, C. M., Ellis, B. A., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science, 321(5894),* 494-495.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why female are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11(5),* 366-371.

Jacobs, E. J., Lanza, S., Osgood, W. D., Eccles, S. J., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73(2),* 509-527.

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. American Psychological Society, 16(3), 175-179.

Keller, J. (2002). Blatant stereotype threat and women's math performance: self handicapping as a strategic means to cope with obtrusive negative performance expectations: Brief report. *Sex Roles, 47,* 193-198.

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

Keller, J. (2007). Effects of single-sex secondary schools on student achievement and attitudes. *Journal of Educational Psychology, 78(5),* 381-395.

Keller, J., & Molix, L. (2008). When women can't do math: The interplay of self-construal, group identification, and stereotypic performance standards. *Journal of Experimental Social Psychology, 44,* 437-444.

Klebanov, K. P., & Brooks-Gunn, J. (1992). Impact of maternal attitudes, girls adjustment and cognitive skills upon academic performance in middle and high school. *Journal of Research on Adolescence, 2(1),* 81-102.

Lee, E. V., & Bryk, S. A. (1986). Effects of single-sex secondary schools on student achievement and attitudes. *Journal of Educational Psychology, 78(5),* 381-395.

Lummis, M., & Stevenson, W. H. (1990). Gender differences in beliefs and achievement: A cross-cultural study. *Developmental Psychology, 26(2),* 254-263.

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

Mael, F., Alonso, A., Gibson, D., Rogers, K., & Smith, M. (2005). Single-sex versus coeducational schooling: A systematic review. *American Institutes for Research*, p. 1 - 148.

Malcolm, G. S. (1986). Women account for half of college enrollment in US., 3 other nations. *Chronicle of Higher Education, 33(2),* 1-42.

Marsh, H. W. (1991). Public, catholic single-sex, and catholic coeducational high schools: Their effects on achievement, affect, and behaviors. *American Journal of Education, 99(3),* 330-356.

Marx, M. D., & Goff, A. P. (2005). Clearing the air: The effect of experimenter race on target's test performance and subjective experience. *British Journal of Social Psychology, 44,* 645-657.

Marx, M.D., & Stapel, A.D. (2006). Distinguishing stereotype threat from priming effects: On the role of the social self and threat-based concerns. *Journal of Personality and Social Psychology, 91(2),* 243-254.

Muzzatti, B., & Agnoli, F. (2007). Gender and mathematics: Attitudes and stereotype threat susceptibility in Italian children. *Developmental Psychology, 43(3),* 747-759.

Onwuegbuzie, J. A. (1995). Statistics test anxiety and female students. *Psychology of Women Quarterly, 19,* 413-418.

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

Pekrun, R., Frenzel, C. A., & Goetz, T. (2007). Girls and mathematics – A "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education, 22(4), 497-514.*

Pezdek, K., Berry, T., & Renno, A. P. (2002). Children's mathematics achievement: The role of parents' perceptions and their involvement in homework. *Journal of Educational Psychology, 94(4),* 771-777.

Pollard, S. D. (1998). "The contexts of single-sex classes." *Separated by Sex: A Critical Look at Single-Sex Education for Girls.* (Washington, D.C.: AAUW Educational Foundation.

Rose, J. A., & Rudolph, D. K. (2006). A review of sex differences in peer relationship processes: Potential trade-offs for the emotional and behavioral development of girls and boys. *Psychological Bulletin, 132(1),* 98-131.

Schmader T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38(2),* 194-201.

Schmader, T., & Johns, M. (2003). Stereotype threat and working memory. *Journal of Personality and Social Psychology, 85*, 440-452.

Schmader, T., Johns, M., & Barquissau, M. (2004). The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles, 50, 835-850.*

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

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

Smith, S. C., & Hung, L. (2008). Stereotype threat: effects on education. *Social Psychology of Education, 11(3),* 243-257.

Smith, L. J., Sansone, C., & White, H. P. (2007). The stereotyped task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology, 99(1),* 99-114.

Sneed, M. (2007, July). Sorting x and y: Single sex schools and classes. Power point presentation presented at the Association for Gender Equity Leadership in Education conference, San Diego, CA.

Spencer, S. J., Steele M. C., & Quinn, D. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 87,* 860-875.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52(6),* 613-629.

Steele, C. M. (1998). Stereotyping threat and its threat is real. *American Psychologist 53(6),* 680-681.

Steele, C. M., Aronson, A. J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797-911.

Steele, C. M., Aronson, A. J. (2004). Stereotype threat does not live by Steele and Aronson (1995) alone. *American Psychologist, 59(1),* 47-48.

Steele, J., James, J., & Barnett, R. C. (2002). Learning in a man's world: Examining the perceptions of undergraduate women in male-dominated academic areas. *Psychology of Women Quarterly, 26,* 38-64.

Sundstrom, A. (2006). Beliefs about perceived competence: A literature review. 55, 15-20.

Taylor, A. J. (1997). Factorial validity of scores on the aiken attitude to mathematics scales. *Educational and Psychological Measurement, 57(1),* 125-130.

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. Educational Psychology Review, 6(1), 49-78.

Wigfield, A., & Eccles, S. J. (1994). Children's competence beliefs, achievement values, and general self-esteem. *Journal of Early Adolescence, 14(2),* 107-138.

Wigfield, A., Harold, D. R., Freedman-Doan, C., Eccles, S. J., Suk Yoon, K., Arbreton, J. A. A., & Blumenfield, C. P. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology, 89(3),* 451-469.

Wolf-Wendel, L. (2008). Single-sex institutions: Historical contribution, characteristics of contemporary women's colleges, and the contemporary importance of women's colleges. Education Encyclopedia, 24. Retrieved September 29, 2008, from education.stateuniversity.com/pages2420/Single-Sexinstitutions.html

Wout, D., Danso, H., Jackson, J., & Spencer, S. (2008). The many faces of stereotype threat: group- and self threat. *Journal of Experimental Social Psychology, 44(3),* 792-799.

Appendices

Appendix A

Please circle and or specify your race/ethnicity identification:

Alaskan/American Indian
Specify:____________

Hawaiian/Pacific Islander
Specify:____________

Asian
Specify:____________

Hispanic/Latino(a)
Specify:____________

Black/African American
Specify:____________

White/Caucasian
Specify:____________

Biracial/Multiracial
Specify:____________

Appendix B

Learn About You…

Please answer the questions below!

Name:

Age:

Grade: 5th or 6th

Please circle your primary race:

Alaskan/American Indian
Hawaiian/Pacific Islander
Asian
Hispanic/Latino
Black/African American
White/Caucasian

Favorite Color:

Did you have boys in your last school? Yes or no

If you did NOT have boys in your last school please write the name of the school you last attended and how many years you attended.

Do you like math? Yes or no

Favorite Subject in School:

Appendix C

Math Test

There are 15 math problems.

Please complete all the problems you can.

Problems not completed do not count against you.

You have 25 minutes to complete.

Please don't guess.

Skip any problems if necessary.

Use your calculator and your mind!

Mathematical Chart

LENGTH

Metric	Customary
1 kilometer = 1000 meters	1 mile = 1760 yards
1 meter = 100 centimeters	1 mile = 5280 feet
1 centimeter = 10 millimeters	1 yard = 3 feet
	1 foot = 12 inches

CAPACITY AND VOLUME

Metric	Customary
1 liter = 1000 milliliters	1 gallon = 4 quarts
	1 gallon = 128 ounces
	1 quart = 2 pints
	1 pint = 2 cups
	1 cup = 8 ounces

MASS AND WEIGHT

Metric	Customary
1 kilogram = 1000 grams	1 ton = 2000 pounds
1 gram = 1000 milligrams	1 pound = 16 ounces

Perimeter	square	$P = 4s$
	rectangle	$P = 2l + 2w$ or $P = 2(l + w)$
Circumference	circle	$C = 2\pi r$ or $C = \pi d$
Area	square	$A = s^2$

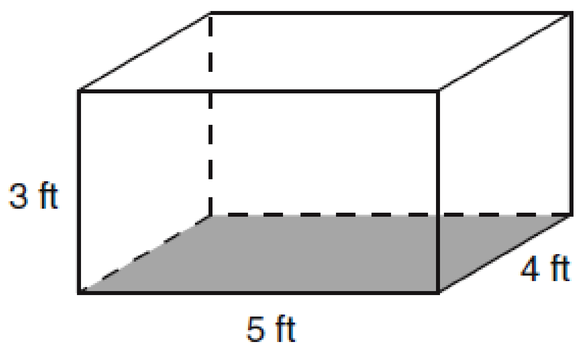
Session A

________ 1) Which digit is in the thousands place in the number 4,861,392?

A 6
B 4
C 1
D Not Here

________ 2) A rectangular rabbit cage is shown below. What is the perimeter of the bottom of the rabbit cage?

F 12 feet
G 16 feet
H 18 feet
J 20 feet



________ *3) Stylists at a hair salon charge $26 for each haircut. If they gave 63 haircuts, how much money did they collect, not including tips?

A $89
B $504
C $1,538
D $1,638

________ 4) Which of these shapes could never have perpendicular lines?

F Square
G Rectangle
H Triangle
J Circle

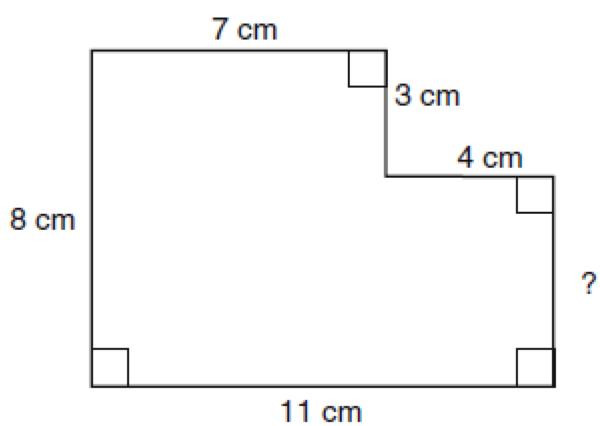
________ 5) According to a report published in 1999, the population of Dallas was 1,063,292. What does the 6 in this number represent?

A Six thousand
B Sixty thousand
C Sixty-three thousand
D Six hundred thousand

_________ 6) On a class field trip, there was 1 adult for every 8 students. If a total of 54 students and adults went on the trip, how many were students?

A 46
B 47
C 48
D 62

_________ 7) The figure below is missing a measurement for one line segment. What is the missing measurement?



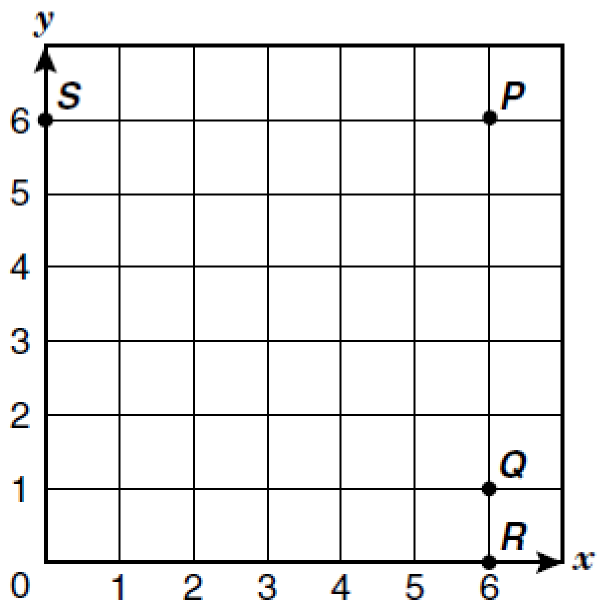
A 1 centimeter
B 4 centimeters
C 5 centimeters
D 8 centimeters

________ 8) Wilma ran 4 miles. She wants to find her running time per mile in minutes. What additional information does she need?

F The number of minutes that she ran
G The number of feet in 4 miles
H The number of laps in 1 mile
J The number of laps that she ran

________ 9) Which point is located at (6, 0)?

F Point *P*
G Point *Q*
H Point *R*
J Point *S*



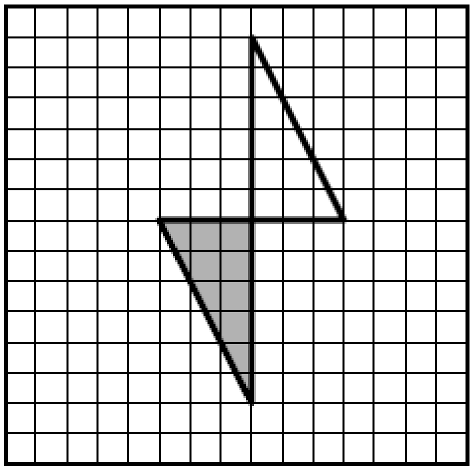
________ 10) Carmen recorded the colors of the first 24 cars that drove by her house Saturday morning. The table shows the data she collected. Which fraction represents the number of black cars that she counted?

A 1/3
B 1/4
C 1/5
D 1/6

Car Colors

Color	White	Black	Red	Blue	Green	Other
Number of Cars	4	6	2	5	4	3

________ 11) Which transformation of the shaded figure is represented in the diagram?



A Reflection
B Translation
C Rotation
D Not Here

________ 12) Joey has 8 books. Roberto has twice as many books as Joey has. How many books does Roberto have?

________ 13) Caleb and his brother collect seashells. Caleb has 468 seashells, and his brother has 263. How many more seashells do they need to collect in order to reach a goal of 750 seashells?

________ 14) Mr. Perkins needs 16 ounces of milk for a recipe. How many cups of milk does he need for the recipe?

Note:
1 liter = 1000 milliliters
1 gallon = 4 quarts
1 gallon = 128 ounces
1 quart = 2 pints
1 pint = 2 cups
1 cup = 8 ounces

F 2 c
G 4 c
H 8 c
J Not Here

________ 15) Dora's family bought a bag of 24 oranges. There are 6 people in Dora's family. If they ate of 3/8 of the oranges, what fraction of the oranges remained?

A 9/8
B 5/8
C 3/14
D 3/48

Session B

_______ *1). Stylists at a hair salon charge $26 for each haircut. If they gave 63 haircuts, how much money did they collect, not including tips?

A. $89
B. $504
C. $1,538
D. $1,638

________2) Some of the greatest long-jump distances by Olympic Athletes are listed in the table below. According to this table which year, in which year was the greatest long-jump distance recorded?

Long-Jump Distances

Year	Distance (meters)
1968	8.90
1976	8.35
1988	8.72
1992	8.67

A. 1968
B. 1976
C. 1988
D. 1992

_________3) Which is a prime factor of the composite number 18?

F. 3
G. 5
H. 6
J. 9

________4) How many millimeters are equivalent to 400 centimeters?

F. 0.4 mm
G. 4 mm
H. 40 mm
J. Not Here

________5) The table shows the number of tickets sold at the first 5 football games. About how many tickets were sold for the first 5 games?

Ticket Sales

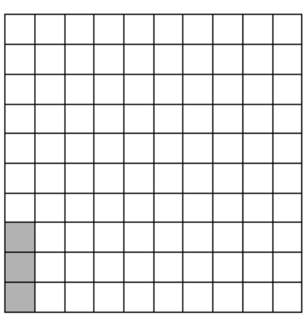
Game	Number Sold
First	263
Second	198
Third	303
Fourth	279
Fifth	234

A. 800
B. 1,000
C. 1,300
D. 1,500

________6). Marcus and Joe are placing books on the library shelves. They have completed of 2/3 of the job so far. Which fraction is equivalent to 2/3?

F. 2/6
G. 2/4
H. 4/6
J. 3/2

______7). What part of the model is shaded?



A 0.003
B 0.03
C 0.3
D 3.0

_______8). Alaska, the largest state in the United States, has an area of 656,424 square miles. Rhode Island, the smallest state, has an area of 1,545 square miles. What is the difference between the areas of these two states?

F. 501,924 sq mi
G. 654,879 sq mi
H. 655,879 sq mi
J. 657,969 sq mi

_________9). Max is packing books into boxes. Each box can hold 12 books. Which number sentence can be used to find the total number of boxes that he needs in order to pack 84 books?

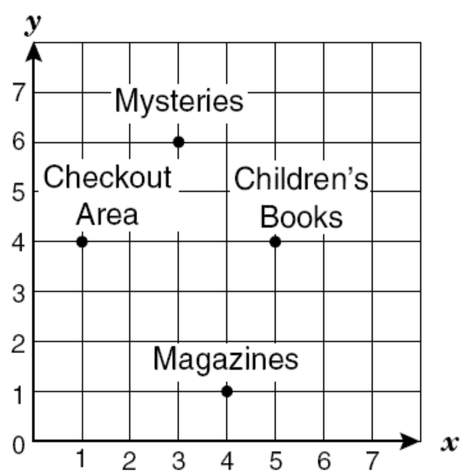
F. 84 ÷ 12 = X
G. 84 - X =12
H. 84 + 12 = X
J. X ÷ 12 = 84

_________10). Read the number 2,309,758,011. Write the number in the place-value chart below.

Billions	Hundred Millions	Ten Millions	Millions	Hundred Thousands	Ten Thousands	Thousands	Hundreds	Tens	Ones

________11). The graph shows some areas of the public library. Which ordered pair best represents the point on the graph labeled "Magazines"?

A (1, 4)
B (3, 6)
C (4, 1)
D (5, 4)



________12). Olivia bought some candy for $0.58. She received $0.42 in change. What is the least number of coins she could have received?

A 4
B 5
C 6
D 7

________13). A recipe for pancakes requires 3 eggs and makes 12 pancakes. What is the ratio of eggs to pancakes?

A 12:3
B 1:4
C 3:1
D 1:3

________14). Karen ordered 15 cases of cookies to sell for her soccer team. Each case had 288 cookies in it. How many cookies did she order in all?

A 1,728
B 4,280
C 4,320
D Not Here

________15). Fabian has organized 3/5 of his baseball card collection. Which decimal represents the fraction of Fabian's collection he has organized?

F 0.4
G 0.6
H 0.35
J 0.12

Session C

______1). Use the place-value chart to order these decimals from greatest to least.

1.19 3.417 3.6 1.1

Ones	.	Tenths
	.	
	.	
	.	
	.	

_____2). At a community center there are a total of 11 tables in the dining room. Six of the tables seat 4 people each. Five of the tables seat 8 people each. What is the maximum number of people who can sit at the tables in the dining room?

F 40
G 12
H 22
J 64

______3). Each week Mary spends $3.95 on newspapers. So far this year Mary has spent $97.50 on newspapers. Which is the best estimate of the number of weeks she has bought newspapers?

A 20 weeks
B 25 weeks
C 40 weeks
D 50 weeks

_______4). Paul made 11 of the 20 shots he took at hockey practice yesterday. What percent of his shots did he make?

F 20%
G 11%
H 45%
J 55%

______5). How many millimeters are equivalent to 800 centimeters?

F 0.8 mm
G 8 mm
H 80 mm
J Not Here

________6). The table shows the cost of different numbers of bus tickets. Which expression can be used to find the cost of 8 tickets?

Bus Tickets

Number of Tickets	Cost of Tickets
2	$4
4	$8
6	$12
10	$20

A $2.00 + 8
B $4.00 + 8
C $1.50 x 8
D $2.00 x 8

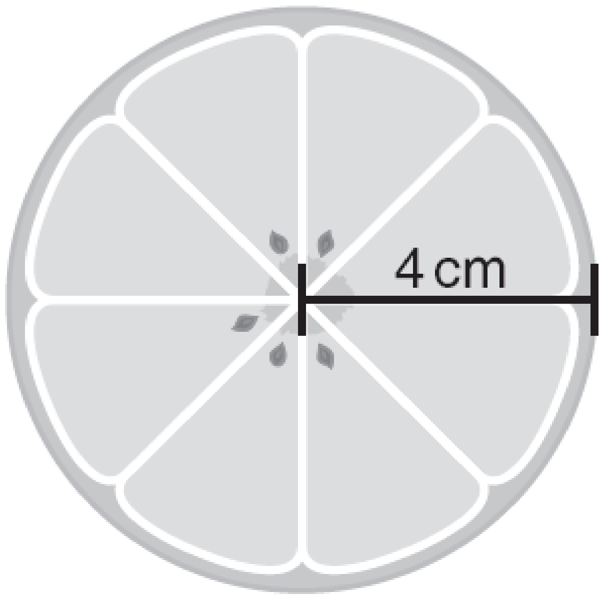
_______7). Mr. Cartwright bought 18 computer books for his bookstore. If he paid $24.95 for each book, about how much did Mr. Cartwright pay for the books?

A $40
B $600
C $500
D $300

_____8). Mrs. Sandoval has 60 folders, 45 pairs of scissors, and 30 rulers. What is the greatest common factor Mrs. Sandoval can use to divide the school supplies into equal groups?

A 3
B 5
C 10
D 15

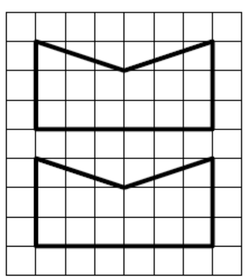
______9). Rosa sliced an orange into circular pieces to put into a bowl of punch. The piece shown below had a radius of 4 centimeters. Which expression can be used to find the approximate circumference of this Orange?



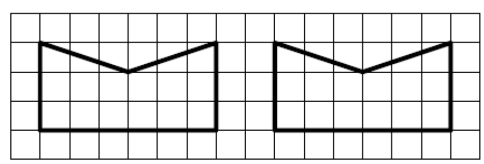
A 2(4)
B π (4)
C 2(π)(8)
D 2(π)(4)

______10). Which of these does **NOT** show a reflection?

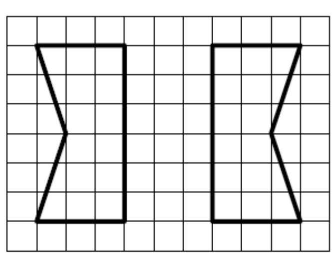
A



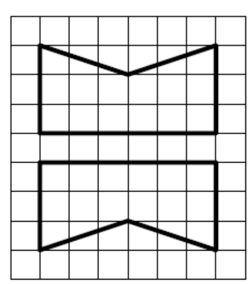
C



B



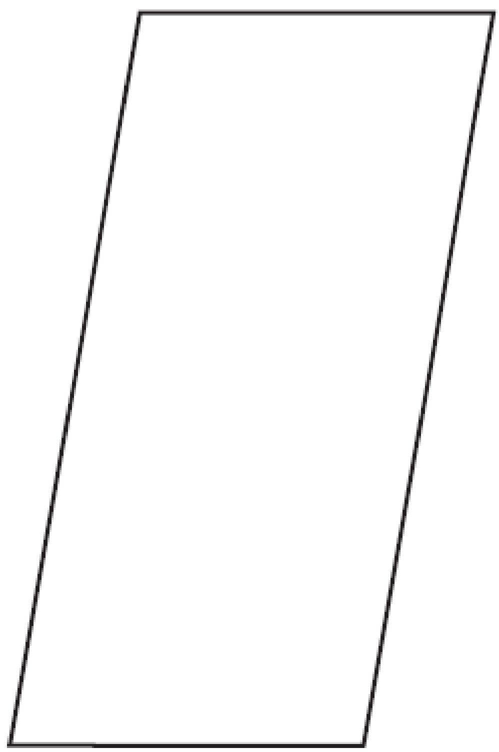
D



_____11). The total length of all the songs on a CD Mohammed bought is about 80 minutes. Each song is between 4 and 6 minutes long. Which is a reasonable number of songs that could be on the CD?

A 10
B 40
C 74
D 16

______12). Look at the parallelogram shown below. Which of the following could be the measures of the angles of the parallelogram?



F 120°, 60°, 120°, 120°
G 80°, 100°, 80°, 100°
H 90°, 90°, 120°, 60°
J 100°, 90°, 80°, 90°

______13). Jeremy received $70 as a gift. He wanted to use the money to go to the movies and to buy a book. He wanted to save the money he had left. Which is the correct order of steps to find the amount of money Jeremy would have left to save?

Step K: Find the sum of the costs of the movie and the book.
Step L: Find the difference between $70 and the sum of the costs of the movie and the book.
Step M: Identify the cost of the movie and the cost of the book.

A L, K, M
B M, K, L
C L, M, K
D K, L, M

______*14). Stylists at a hair salon charge $26 for each haircut. If they gave 63 haircuts, how much money did they collect, not including tips?

A. $89
B. $504
C. $1,538
D. $1,638

______15). The ratio of women to men in a local book club is 7 to 3. Which combination of women and men could the club have?

A 21 women and 9 men
B 35 women and 50 men
C 14 women and 9 men
D 21 women and 15 men

Appendix D

Questions for You Survey

Section A. Read each statement carefully and decide how you feel right now. Then put an X in the circle in front of the word or phrase which best describes how you feel. There are no right or wrong answers. Do not spend too much time on any one statement. Remember, find the word or phrase which best describes how you feel right now, at this very moment.

1.	I feel…	O very calm	O calm	O not calm
2.	I feel…	O very upset	O upset	O not upset
3.	I feel…	O very pleasant	O pleasant	O not pleasant
4.	I feel…	O very nervous	O nervous	O not nervous
5.	I feel…	O very jittery	O jittery	O not jittery
6.	I feel…	O very rested	O rested	O not rested
7.	I feel…	O very scared	O scared	O not scared
8.	I feel…	O very relaxed	O relaxed	O not relaxed
9.	I feel…	O very worried	O worried	O not worried
10.	I feel…	O very satisfied	O satisfied	O not satisfied
11.	I feel…	O very frightened	O frightened	O not frightened
12.	I feel…	O very happy	O happy	O not happy
13.	I feel…	O very sure	O sure	O not sure
14.	I feel…	O very good	O good	O not good
15.	I feel…	O very troubled	O troubled	O not troubled
16.	I feel…	O very bothered	O bothered	O not bothered
17.	I feel…	O very nice	O nice	O not nice
18.	I feel…	O very terrified	O terrified	O not terrified
19.	I feel…	O very mixed- up	O mixed up	O not mixed-up
20.	I feel…	O very cheerful	O cheerful	O not cheerful

Continue to Next Page.

Section B. Please use the following scale to answer the 4 questions below.

1 = Strongly Disagree
2 = Disagree
3 = Somewhat Disagree
4 = Neutral
5 = Somewhat Agree
6 = Agree
7 = Strongly Agree

_______1) I worry that my ability to perform well on math tests is affected by my being a girl.

_______2) I worry that if I perform poorly on a test, people will think my poor performance is due to me being a girl.

_______3) I worry that people's evaluations of me will be affected by my being a girl.

_______4) I worry that, because I know the negative stereotype about girls and mathematic achievement, my anxiety about proving the stereotype will negatively influence my math performance.

Appendix E

Thoughts Survey

Please answer the following questions using the scale below. Thanks.

1 = strongly disagree
2 = disagree
3 = neutral
4 = Agree
5 = Strongly Agree

1. __________I often think of myself as a girl in school.
2. __________My being a girl is very important to me.
3. __________I am proud to be a girl.
4. __________My being a girl is important to my academic success.
5. __________I often think of myself as a girl in class.
6. __________My being a girl impacts my academic success.

Notes

1. Stereotype Threat Theory is also used to explain and describe the role of negative stereotypes in other academic domains, and in areas outside of academia such as sports. Furthermore, the theory has been applied to negative stereotypes affecting other social groups such as African-Americans. See Wheeler & Petty, 2001 for a comprehensive list of stereotype threat studies.
2. The term *performance* refers not only to academic, but applies to other domains such as performance in sports.
3. The work of Neuville and Croizet (2007) and Huguet and Regner, (2007) examined stereotype threat in same sex versus mixed sex settings, but only in the laboratory.
4. The study was longitudinal to account for any variance in the girls' gender (dis)identification as a result of the fact that majority of the participants had previously attended co-ed schools and were transitioning into a single-sex environment. As discussed, gender identity (social identity) can be cued by the gender composition of an environment. Session A provides a baseline for the identification of the girls in the single-sex environment, while session B and C indicate the changes (if any) in the gender identification of the students. Moreover, participants in the stereotype threat condition were prompted to circle their gender prior to completing the math test, and thus, to examine their gender identification at points B and C is critical in examining how the setting and the inducement of stereotype threat may interact.
5. Middle school girls were used for three main reasons highlighted within previous research. The first is that middle school is when gender differences in math emerge including deficits in performance, interest and attitudes. Second, the differentiation of the

sexes or gender identity becomes more salient beginning in middle school. Consequently, gender stereotypes, gender-appropriate expectations and appropriated gendered behavior are formed (Martin & Little, 1990; McKown & Weinstein, 2003) illustrated in research showing that girls in particular place a greater emphasis on success based on their achievement in the appropriate gendered domains in middle school (Klebanov & Brooks-Gunn, 1992). Third, middle school is the age when children become aware of stereotypes and thus, participants in this study should be aware of the negative stereotype regarding girls' and math. Therefore, this literature suggests that middle school girls would be aware of the negative stereotype pertaining to math, be developing a sense of their gender identity and awareness of the negative stereotype, and consequently, stereotype threat could be induced and examined in an experiment. Furthermore, previous stereotype threat studies examining stereotype threat and children have used middle school aged girls (Frome & Eccles, 1998; Keller, 2002; Keller & Dauenheimer, 2003). Studies examining middle school girls have found that girls underperform in stereotype threat conditions.

6. Racial/Ethnic Distribution of Participants: 2 Alaskan/American Indian, 6 Asian, 5 Hispanic/Latino(a), 10 Black/African American, 34 White/Caucasian, 7 Biracial/Multiracial and 1 Other and 8 participants information missing.
7. The model, based on the work of Steele, Spencer and Aronson (2002) consists of (primarily in laboratory settings), a situation in which one positively stereotyped social group and one negatively stereotyped group are required to perform a specific task (Cullens, Water & Sackett, 2006; Muzzatti & Agnoli, 2007).
8. The term *Students* refers to both the participants and non-participants in the classroom during the study.

9. The colors of the materials used in the study included, red, yellow, purple, pink, salmon and blue.
10. Although other studies have used explicit written or oral prompts explicitly indicating (gender or race) “performance differences” on the given task (Marx & Goff, 2005), this procedure would interfere with our analysis of the girls’ (dis)identification with their gender. Thus, the *threat* is induced by the awareness of negative stereotype(s) and is consistent with stereotype threat research as individuals in stereotype threat situations do not have to internalize the stereotype or believe in the stereotype, but only be aware of the negative stereotype for it to *threaten* them (Marx & Stapel, 2006; Shih, Ambady & Pittinsky, 1999; Smith, Sansone & White, 2007; Steele, 1997; Wheeler, Jarvis & Petty, 2001).
11. Accuracy scores have been used within numerous stereotype threat studies to evaluate math performance (Ben-Zeev &Inzlicht, 2000; Keller, 2002; Keller, 2007; Schamder, 2002; Steele & Aronson, 1995).
12. To measure the psychological presence of stereotype threat within these experiments, many researchers use Anxiety scales or Math Anxiety Scales (Steele & Aronson, 1995). In general females report higher anxiety levels than males in completing math tasks and this anxiety has been associated with women’s substandard performance in mathematics (Ashcraft & Faust, 1994; Chipman, Krantz & Silver, 1992; Pekrun, Frenzel & Goetz, 2007), including women in stereotype threat conditions (Keller & Dauenheimer, 2003; O’Brien & Crandall, 2003; Onwuegbuzie, 1995; Pekrun et. al., 2007; Spencer, Steele & Quinn, 1999; Steele & Aronson, 1995) suggesting that anxiety may contribute to the participants’ suboptimal performance

13. Anxiety scores are based on a range of values from 1 to 2 with scores closer to 1 indicating low anxiety and scores closer to 2 indicating high anxiety.

14. Participants in stereotype threat condition correctly completed 61% of the math problems while participants in the non-stereotype threat condition correctly completed 58% of the math problems, although the difference was not significantly significant F (1, .057) = .819, $p < .05$. This data is not in the expected direction or consistent with the previous work of stereotype threat that found participants in stereotype threat conditions had lower accuracy scores (worse math performance) than participants in non-stereotype threat conditions (Huguet & Regner, 2007; Onwuegbuzie, 1995; Schmader & Johns, 2003; Spencer, Steele & Quinn, 1999).

Figure Captions

Figure 1. Mean accuracy scores of participants based on overall gender identification, strong, weak or changed.

Figure 2. Mean accuracy scores of participants' at Session A and B based on condition, stereotype threat or non-stereotype threat.

Figure 3. Frequency of participants overall gender identification at Session A and B.

Figure 1.

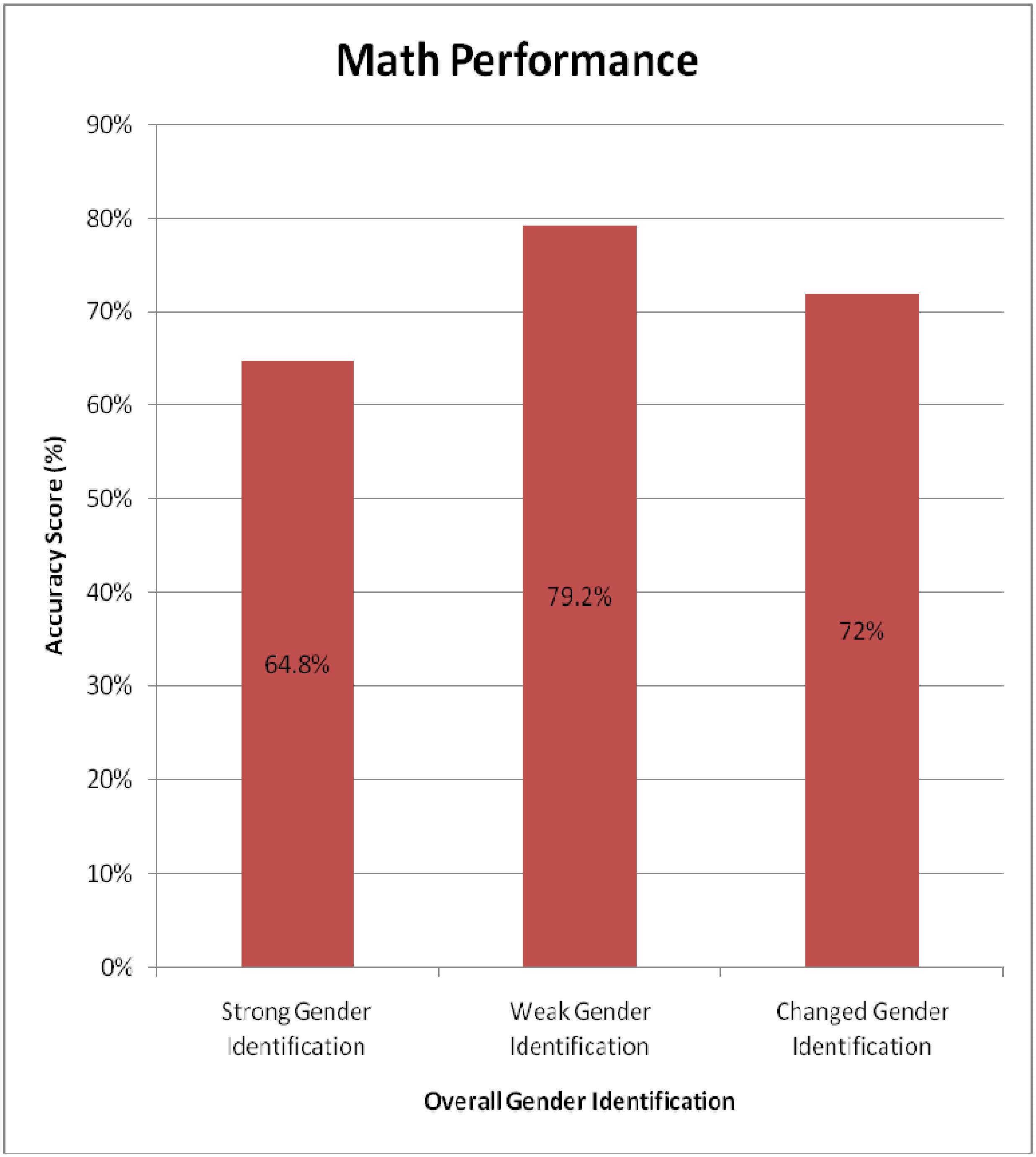

Math Performance
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Accuracy Score (%)
64.8%
79.2%
72%
Strong Gender Identification
Weak Gender Identification
Changed Gender Identification
Overall Gender Identification

Figure 2.

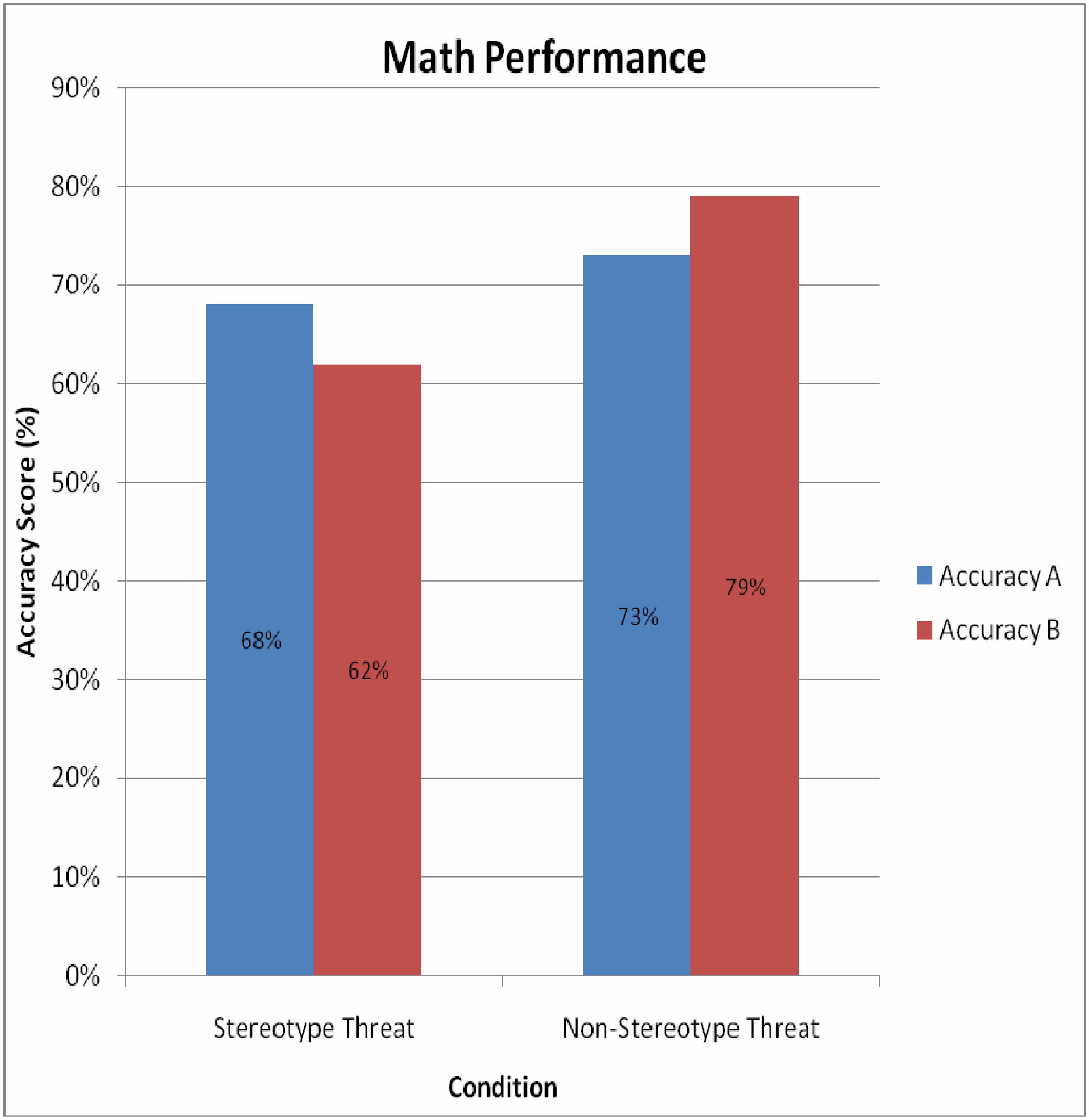
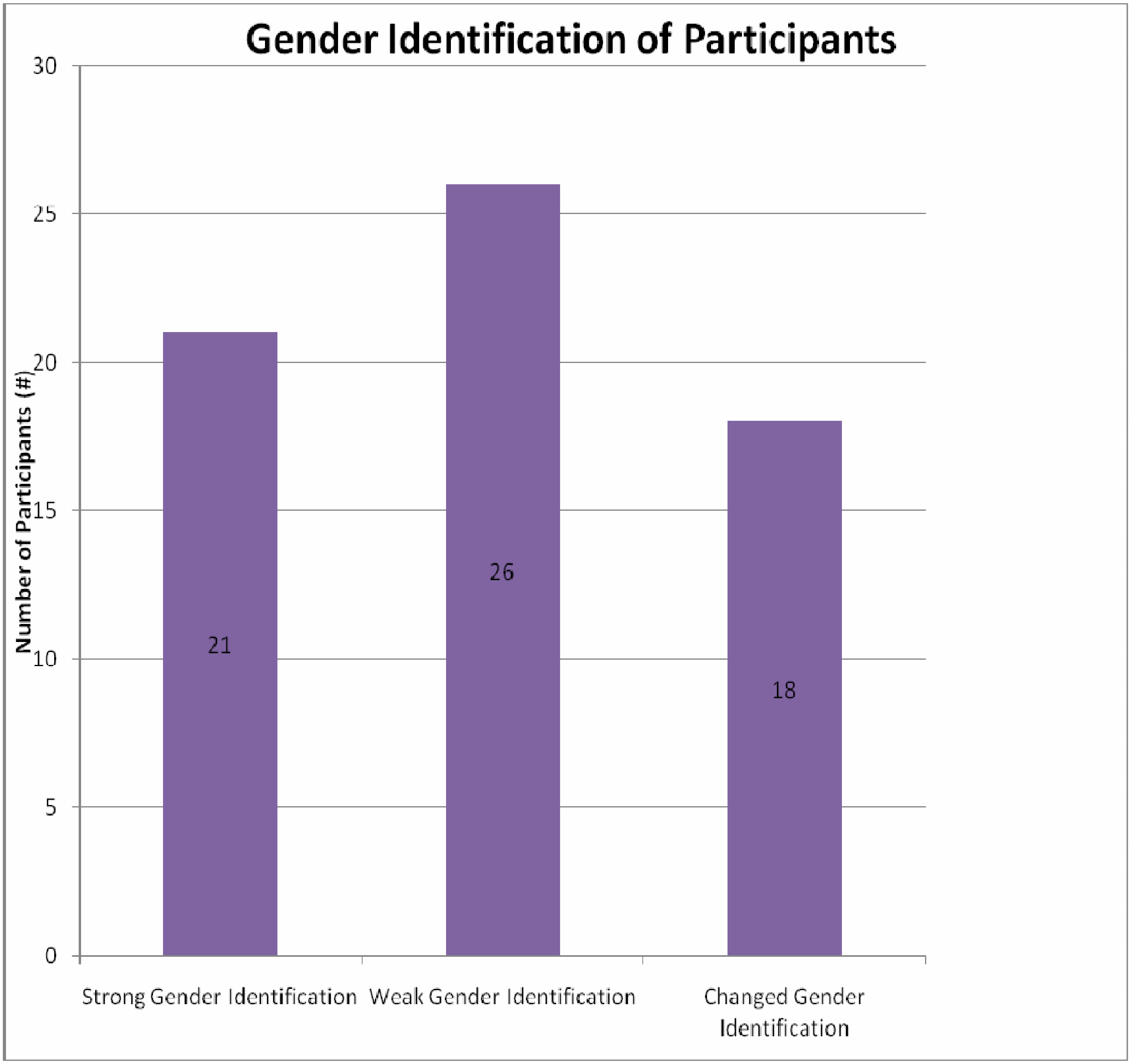

Math Performance
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Accuracy Score (%)
68%
62%
73%
79%
Accuracy A
Accuracy B
Stereotype Threat
Non-Stereotype Threat
Condition

Figure 3.


Gender Identification of Participants
Number of Participants (#)
30
25
20
15
10
5
0
21
26
18
Strong Gender Identification
Weak Gender Identification
Changed Gender Identification