THAT WAS HARD! EXAMINING THE EFFECTS OF TEST INSTRUCTIONS AND CONTENT ON WOMEN'S MATHEMATICS PERFORMANCE UNDER STEREOTYPE THREAT

by

DANA PRESTWOOD GRESKY

Bachelor of Science, 2000
Palm Beach Atlantic University
West Palm Beach, Florida

Master of Science, 2003
Texas Christian University
Fort Worth, Texas

Submitted to the Graduate Faculty of the
College of Science and Engineering
Texas Christian University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

December 2006

ACKNOWLEDGMENTS

It is exciting to be able to write an acknowledgments page. First, it is important to express my sincere thanks to my major professor, Dr. Charles Lord. He has always expressed confidence in my ability to achieve, and I am truly thankful.

I am also grateful for the counsel of my dissertation committee: Dr. Charles Bond, Dr. David Cross, Dr. Donald Dansereau, and Dr. Christie Napa-Scollon.

I met my husband just three weeks before defending my master's thesis. At the time, I simply told him that I liked him, but I could not see him for a month! His positive response made me sure he was a winner. Since then, he has continued to welcome my pursuit of education, and has loved me despite so many omitted dinners and paper-filled rooms.

I thank my parents for congratulating me for every milestone. They have encouraged me to continue with my education by posting since I was a child by posting Colossians 3:23 in my bedroom as a child: "Whatsoever you do, do it heartily as unto the Lord, and not unto men".

Finally, I thank my dear friend and colleague, Dr. Laura Ten Eyck, for all of the emotional and wonderful times we have experienced together: crying during our first year, sleeping in the office, all of the self-disclosure, and watching the other find her soul mate…

TABLE OF CONTENTS

Acknowledgements……………………………………………………………………………ii

List of Figures…………………………………………………………………………...iv

List of Tables…………………………………………………………………………....v

I. Introduction…...………………………………………………..…………..…...1

II. Pretest……………………………………………………………………….21

III. Experiment 1……………………………………………………………..22

IV. Experiment 2……………………………………………………………..49

V. General Discussion…..……………………………………………………..66

VI. Footnote……………………………………………………………………70

Appendices……………………………………………………………………………...71

References……………………………………………………………………………..125

Vita

Abstract

LIST OF FIGURES

1. Intended level of Difficulty for Each Item (Experiment 1)……………..…………..27

2. Actual Difficulty Levels for Each Item as Determined During Pretesting (Experiment 1)……………………………………………………………………………..28

3. Average Difficulty Ratings for Each Math Item by Men and Women in Both Conditions (Experiment 1)…………………………………………………………...30

4. Average Difficulty Ratings for Each Math Item by Women Given Punishment and Reward Instructions (Experiment 1)………………………………...……………..31

5. Average Percent Correct for Each Math Item by Men and Women (Experiment 1)……………………………………………………...……………………...35

6. Average Percent Correct for Each Math Item by Participants in Each Condition (Experiment 1)……………………………………………………………….……36

7. Percentage of Correct Items in Each Section of the Math Test by Women (Experiment 1)………………………………………………………………………39

8. Mean Response Times in Seconds on Target and Non-target Items for Men and Women in Each Condition (Experiment 2)…………………………………..…...58

LIST OF TABLES

1. Mean Difficulty Ratings of Ambiguous and Manipulation Items by Participants in each Condition (Experiment 1) ……………………..…………………………….32
2. Mean Difficulty Ratings of Ambiguous and Manipulation Items by Women in Each Condition (Experiment 1)…………………………………………………………34
3. Mean Percentage Correct Responses for Items 10-14 and 19-22 for Men and Women (Experiment 1)……………………………………………………….…………38
4. Means, Standard Deviations, and Correlations of Target Math Items and Gender, Condition, and Domain Identification (Experiment 1)………………………………41
5. Linear Regression Predicting Percent Correct Scores on Items 10-14 from Gender, Condition, Domain Identification, and Their Interactions (Experiment 1)………….42
6. Linear Regression Predicting Percent Correct Scores on Items 19-22 from Gender, Condition, Domain Identification, and Their Interactions (Experiment 1)………….43
7. Means, Standard Deviations, and Correlations of Target Math Items, Condition, and Domain Identification for Women Only (Experiment 1)……………………………45
8. Linear Regression Predicting Percent Correct Scores on Items 10-14 from Condition, Domain Identification, and Their Interactions for Women Only (Experiment 1)…...46
9. Linear Regression Predicting Percent Correct Scores on Target Items from Condition, Domain Identification, and Their Interactions (Experiment 1)……………………...47
10. Total Number of Correct, Incorrect, and Impossible Answers on Each of the Shared Math Items (Experiment 2)………………………………………………………..53
11. Mean Number of Non-Target Math Items Correct by Women and Men (Experiment 2)…………………………………………………………………………………..54

12. Means, Standard Deviations, and Correlations of Target Math Items and Gender, Condition, and Domain Identification (Experiment 2)……………………………56

13. Means, Standard Deviations, and Correlations of Target Math Items and Gender, Condition, and Domain Identification (Experiment 2)……………………………60

14. Linear Regression Predicting Percent Correct Scores on Target Items from Gender, Condition, Domain Identification, and Their Interactions (Experiment 2)………….61

15. Means, Standard Deviations, and Correlations of Target Math Items, Condition, and Domain Identification for Women Only (Experiment 2)……………………………63

16. Linear Regression Predicting Percent Correct Scores on Target Items from Condition, Domain Identification, and Their Interactions (Experiment 2)……………………...64

That Was Hard! Examining the Effects of Test Instructions and Content on Women's Mathematics Performance Under Stereotype Threat

Why can't we seem to get it together? A glance at the national average math scores on the Scholastic Assessment Test, the National Longitudinal Study of the High School Class of 1972, and the National Study of Youth all present data that favor males (Hedges & Nowell, 1995). This is no new phenomenon; women have earned lower scores than males for the last 40 years (The College Board, 2003).

Our culture is ripe with stereotypical examples of the link between femininity and poor math skills. For example, in 1994 Mattel released a Teen Talk Barbie that said 'clever' phrases, such as "I love shopping," and "Math is hard." It is possible that stereotypical representations like these may reflect genetic differences between men and women, as some researchers have proposed (Benbow & Stanley, 1980). A second possibility is that rather than recorded differences in mathematics leading to stereotypes, it is possible that they are the result of the stereotypes.

A meta-analysis of gender differences in mathematics revealed that, during elementary school years, girls initially earn higher scores on standardized mathematics tests than boys (Hyde, Fennema, & Lamon, 1990). As the children age, boys begin to earn higher scores in high school and the trend continues into adulthood. Other researchers have demonstrated a link between the emergence of stereotype awareness among minority group members (i.e. African Americans, women) and a decrease in performance related to the stereotype (McKown & Weinstein, 2003).

How exactly do stereotypes led to differences in performance? For African Americans, stereotypes have lead to racial stratification in the United States, resulting in different educational practices for ethnic groups (Ogbu, 1994). Different educational practices can ultimately lead to marked differences in academic ability between ethnic groups. In addition, African Americans also use stereotypes to define what it means to be Black in America. The stereotypical representation of Whites includes academic achievement, a topic that is often left out of Black representations. In an effort to adhere to a more ethnically-consistent representation, African Americans often choose a route other than academics. Similar experiences can describe the woman's perspective. Women majoring in math-related fields who agree with the stereotype that women are worse than men at math often choose a career route outside of mathematics (Schmader, Johns, & Barquissau, 2004). Additionally, when women are reminded of well-known stereotypes, they tend to make excuses for their performance by using self-handicapping strategies, like calling the test unfair or blaming their test performance on stress (Keller, 2002). Unfortunately, it takes very little to remind women of the negative stereotype regarding their group. In fact, completing a math test in small groups in which women are the minority is sufficient to bring to mind negative stereotypes about women's ability (Inzlicht & Ben-Zeev, 2000).

Being reminded of stereotypes about one's group is an important factor in explaining group differences in performance. Researchers in the field of stereotype threat have thoroughly described how this reminder, along with other factors can debilitate performance across many domains. Stereotype threat has been implicated as one of the explanations for the long-standing performance gap between males and females on standardized tests, such as the Scholastic Assessment Test (SAT), and the gap between minority and Caucasian students on tests of verbal ability (Steele, Spencer, & Aronson, 2002).

Stereotype Threat Characteristics

Performance debilitation. Stereotype threat is a performance debilitating effect that occurs when people are made aware of a negative stereotype concerning their performance (see Steele, 1997 for a review). Even before a test begins, people under stereotype threat maintain lower expectations about their performance on an upcoming test than people who are not under stereotype threat (Stangor, Carr, & Kiang, 1998). After a test begins, a description of a test, such as one that is "diagnostic of their abilities," may lead to impaired performance for stereotype threatened individuals (i.e. Steele & Aronson, 1995, Experiment 2).

Situational threat. In order to experience stereotype threat, situational cues must suggest the possible application of a stereotype. Researchers have been able to create situational cues in a number of ways. For instance, in one study experimenters were able to elicit stereotype threat effects by telling a group of participants that their scores on a mathematics test would be announced to the group (Inzlicht & Ben-Zeev, 2000, 2003). The performance of women completing a difficult math test in a same-sex group was not inhibited by the announcement; however, women completing the math test in a mixed-sex group were inhibited. In fact, their performance deficits were proportional to the number of males present.

Researchers have also induced stereotype threat simply by making salient the identity that is negatively stereotyped (Ambady, Shih, Kim, & Pittinsky, 2001; Bosson, Haymovitz, & Pinel, 2004; Shih, Pittinsky, & Ambady, 1999; Walsh, Hickey, & Duffy, 1999). For instance, in one study researchers asked Asian American girls (kindergarten through eighth grade) questions about their ethnicity or about their gender before completing a math test.

They found that, among middle school girls, making gender salient was sufficient to lower math test scores (Ambady et al., 2001).

Research on stereotype threat also employs an explicit mention of stereotypes concerning the performance of certain groups of individuals (Arndt, Greenberg, Schimel, Pyszczynski, & Solomon, 2002; Aronson, Lustina, Good, Keough, Steele, & Brown, 1999; Cadinu, Maass, Frigerio, Impagliazzo, Latinotti, 2003; Franz, Cuddy, Burnett, Ray, & Hart, 2004; Keller, 2002; Gresky, Ten Eyck, Lord, & McIntyre, 2005; Leyens, Désert, Croizet, & Darcis, 2000; McIntyre, Paulson, & Lord, 2003; Smith & White, 2002). For instance, experimenters were able to induce stereotype threat effects by explicitly stating that "men are poorer performers than women in affective processing tasks" (Leyens, et al., 2000, p. 1189).

Other studies of stereotype threat include a stereotype threat induction that is more subtle (Blascovich, Spencer, Quinn, & Steele, 2001; Brown & Pinel, 2003; Davies, Spencer, Quinn, & Gerhardstein, 2002; Josephs, Newman, Brown, & Beer, 2003; Keller & Dauenheimer, 2003; Oswald & Harvey, 2000; Schmader & Johns, 2003; Smith & White, 2002; Spencer, Steele, & Quinn, 1999; Stone, Lynch, Sjomeling, & Darley, 1999). The experimenter might mention that group differences may exist, without explicitly mentioning the direction of the differences. When men and women are told that researchers have found differences between men and women, participants typically interpret this statement as "men are better than women" (Spencer et al., 1999).

Researchers might simply indicate that they are interested in the performance of the participants, without making specific mention of any stereotypes or expected group differences (Brown & Josephs, 1999, Study 1; Croizet & Claire, 1998; Gonzales, Blanton, & Williams, 2002; McFarland, Lev-Arey, & Ziegert, 2003). In this design, the experimenter may mention that the test is an indicator of the participants' 'true' abilities. Women and minorities performing under these conditions often fall victim to typical stereotype threat effects.

Many groups are affected. Although it may typically be thought of as an effect that harms only women's math performance and African American's verbal skills performance, stereotype threat can affect performance in other domains and for other groups. For instance, researchers framed a golf putting task as indicative of sports intelligence, a label typically applied to the skills of White athletes (Stone et al., 1999). With that test description, African Americans performed worse than control participants. In a second experiment, the researchers framed the same golf putting task as indicative of natural athletic ability, a label typically applied to the skills of Black athletes. The new description harmed the performance of White athletes, causing them to execute more swings to sink a putt.

Researchers have been able to demonstrate that the commonly held stereotype, "men are not sensitive to emotion," can lead to stereotype threat effects (Leyens, et al., 2000). When males were informed that they tend to do worse on affective processing tasks, they not only perform worse on the task than other groups, but they also seem to make more 'false positive' mistakes that suggest they are trying to be more "in tune" with their feelings.

People who hold a particular socioeconomic status (SES) can serve as a stereotyped reference group (Croizet & Claire, 1998). When low-SES participants are reminded about their status they had poorer performance than high-SES participants.

Aging is another topic about which we hold many stereotypes. Older adults have demonstrated typical stereotype threat effects on memory tasks (Hess & Hinson, 2006). After reading an article that supported stereotypical views of memory and aging, older adults took part in a short memory task. They performed worse on the memory task than adults who had previously read an article that contradicted the stereotypical view of aging and memory.

Moderators of Stereotype Threat

Task difficulty. Imagine a college-aged man and a college-aged woman both taking the same difficult math test. They are both likely to experience frustration and anxiety while taking the test. But women have an added burden: a fear that their behavior will confirm the stereotype that women are bad at math. When a woman is completing an easy test, she should not experience frustration or anxiety and will not fear confirming the stereotype about inferior math skills. In one study, researchers gave men and women one of two types of tests: an easy test or a difficult test (Spencer et al., 1999, Study 1). They found typical stereotype threat effects only for the difficult test. Women's performance was no different from men's on the easy test, but women performed significantly worse than men on the difficult test.

Diagnosticity. In order for performance to be affected by stereotype threat, an individual must feel as though his or her performance will indicate success or failure in a relevant domain (Steele, 1997). Thus, stereotype threat effects require a test description that includes the possibility that scores on the test will indicate support of a stereotype. In one study, researchers described a verbal skills test to students with general baccalaureate and technological baccalaureate degrees as being either a test of intellectual ability (diagnostic) or as a test of a cognitive psychology hypothesis (non-diagnostic; Croizet, Dutrevis, & Desert, 2002). In the diagnostic condition, the students with technological baccalaureates were reminded of a stereotype about their group that they are poor performers at intellectual tasks. As a result, their performance on the task suffered. The performance of students with technological baccalaureates in the non-diagnostic condition, however, was unaffected. Other studies have also demonstrated the link between test diagnosticity and stereotype threat effects (Mayer & Hanges, 2003; McKown & Weinstein, 2003; Steele & Aronson, 1995, Study 1 & 2; Stone et al., 1999).

Stereotype relevance. Stereotype threat researchers have clearly indicated that the debilitating effects that have been found are not limited to individuals who consider the stereotypes to be true or predictive of their performance (Steele, 1997). To clarify, it is not a belief that a stereotype is valid, but rather that a stereotype can apply to an individual, based upon group membership, and that the stereotype conflicts with the individual's personally held expectations concerning his or her performance.

Domain identification. Research on stereotype threat suggests the importance of identifying with the task (Schmader, 2002; Smith & White, 2001, Wheeler & Petty, 2001). Researchers have found that among those highly identified with the performance domain, simply providing a difficult test is sufficient to elicit stereotype threat effects (Spencer et al., 1999, Study 1). Individuals performing within these parameters should be more aware of negative stereotypes applied to their ingroup.

Group identification. In order for a woman to be threatened by a stereotype about women, she must also consider her identity as a woman central to her self-definition. According to social identity theory (see Schmader, 2002), we have social identities which are conceptions of self as a member of different social groups. Each of us is motivated to maintain a positive social identity. Thus, if a woman considers her identity as a woman important, stereotype threat should be a more overwhelming situation than if she considered being a woman less important. Previous research has shown that when women are highly identified with their gender, they are more susceptible to stereotype threat effects on mathematics tests.

Mediators

Stereotype activation. In order for stereotype threat to occur, a group member must be reminded of the relevant stereotype. In one study, African American students under stereotype threat completed more word fragments with words that were consistent with negative stereotypes (Steele & Aronson, 1995). Alone, this finding suggests that stereotypes are activated, but not that they affect performance. In a second study, participants watched television commercials before taking a math test (Davies et al., 2002). In one condition, participants saw commercials depicted women stereotypically; such as a woman drooling over a brownie mix. In a second condition, participants saw commercials that depicted women in a way that was counter to stereotypes; for instance, a woman impressing a man with her knowledge of cars. After watching the commercials, participants completed a lexical-decision task designed to measure stereotype activation. Finally, all of the participants took a math test. The results revealed that women who watched stereotypic depictions of women demonstrated poorer math ability and they brought to mind more negative stereotypes than women who watched non-stereotypic depictions of women. Furthermore, it was found that the stereotypes were responsible for their poor performance on the test.

Performance expectations. One strategy test-takers might employ to combat a disadvantageous context is to change their self-image and lower self-expectations in an effort to save face and perform in a manner consistent with their claims. In one study of stereotype threat, participants completed a spatial abilities task and then received either positive or ambiguous feedback (Stangor et al., 1998). After receiving ambiguous feedback, women provided much lower expectations for a future task they were told was sensitive to gender differences, than for a future task they were told was insensitive to gender difference. When women received positive feedback on the first task, their expectations about the second task were no different whether it was presumed to show gender differences or not. It is possible that lowering self-expectations may act as a buffer against possible embarrassment that arises from exhibiting performance that is inconsistent with one's claims.

Mechanisms

As reviewed above, researchers have thoroughly examined stereotype threat characteristics as well as possible mediators and moderators. One area of stereotype threat that has received less investigation is its underlying mechanisms. How does it work? Why does stereotype threat reduce scores? Two areas have been highlighted to answer these questions: reduced cognitive capacity and anxiety.

Reduced cognitive capacity. To begin, researchers have demonstrated effects that are a likely to result from reduced cognitive capacity. For instance, researchers gave men and women a math test. One version of the math test include word problems, the other version included the numerical equivalent of each of the word problems (Quinn & Spencer, 2001). Women underperformed in relation to men on the word problems, but earned equal scores on the numerical equivalent tests. When their problem solving strategies were examined, it was discovered that women used poorer problem solving skills to complete the word problems, a likely result of reduced cognitive capacity. To clarify these findings, researchers set up an experiment to test the effects of reduced working memory capacity directly (Schmader & Johns, 2003). Women were divided into one of two groups. In the stereotype threat condition, women began in a group made up of two male participants and a male experimenter. They were also told that the researcher was collecting data on men and women. In the control condition, women began in a group made up of women and a female experimenter. They were also told that the researcher was interested in collecting data on college students. After receiving instructions, all of the women in the experiment completed the working memory task alone, and then returned to a larger room where they were joined again by their group for further testing. The next portion of the experiment involved completing a difficult math test. Results indicated that stereotype threat reduced working memory capacity and negatively affected math test scores. Furthermore, the reduction in working memory predicted poorer performance on the test.

Anxiety. Arousal may be a typical response to taking tests, but researchers have found that members of stereotyped groups may experience more arousal than non-stereotyped groups. For instance, African Americans demonstrate a physiological indicator of anxiety, increased blood pressure, when in a stereotype threatening situation, and the increased blood pressure is higher than that of Whites taking the same test (Blascovich et al., 2001). A second group of researchers extended the research on arousal, adding an anxiety hypothesis (O'Brien & Crandall, 2003). By giving men and women math tests of varying difficulty, they demonstrated that anxiety has two results; it can improve women's performance on easy tasks and reduce women's performance on difficult tasks. This finding extends classic research on anxiety to demonstrate its applicability in stereotype threat research (Zajonc, 1965).

Stereotype Threat Alleviation Techniques

Because of the well-documented effects of stereotype mentioned above, researchers have extended their focus to include methods that might reduce the effects of stereotype threat. Because stereotype threat can be activated easily, as in situations that include the mere presence of majority group members, researchers have examined many different methods that might reduce the threat.

Reducing stereotype applicability. Providing stereotype threatened individuals with information that is contrary to stereotypes is one way that threat has been alleviated. In one study, researchers provided information that women perform better than men on logical problems (Cadinu, et al., 2003). By doing so, the researchers were able to reduce the applicability of the stereotype that women are poor at math. After receiving the information, women not only had improved performance expectations but improved performance on a math test.

Teaching stereotype threat. One practical approach to reducing stereotype threat involves teaching women about the phenomenon. It follows then that after women are taught about the threat that they may attribute test-taking arousal to stereotype threat, rather than to poor performance. In one study, researchers gave a difficult math test to men and women (Johns, Schmader, Martens, 2005). In the first condition the test was described as a problem-solving task. In the second condition, the test was described as a math test. In the third condition, the test was described as a math test, but women and men were taught about stereotype threat before completing the test. The results indicated that women who learned about the threat earned scores no different from men or from women in the problem-solving group. Women who took that math test without first learning about stereotype threat scored worse than the other groups.

Describing the test as impervious to gender differences. Some researchers have alleviated stereotype threat effects by informing participants that the test that they are about to take does not indicate gender differences. Similarly, researchers might point out that gender differences might exist, but that they have not found differences in their own testing (Sekaquaptewa & Thompson, 2003; Smith & White, 2002; Spencer, et al., 1999, Studies 2 & 3). For instance, in one study, researchers simply told participants that the test that their participants were about to take had not been shown in previous research to produce gender differences in scoring (Smith & White, 2002). In this experiment, women earned math scores equal to those of men. An intervention of this type creates a situation very similar to inducing individuation.

Inducing individuation. One tactic that has successfully been used to alleviate stereotype threat effects has been inducing individuation, such that the person feels less representative of a stereotyped group (Ambady, Paik, Steele, Owen-Smith, Mitchell, 2004). In a relevant study, researchers asked one group of women to describe their favorite book, movie, and so on, while the other group was to spend the same amount of time answering questions about lions. Women who answered questions about their favorite things scored higher on a subsequent math test than did women who answered questions about lions, presumably because describing favorites helps a person feel more like an individual and less like a member of a (stereotyped) group.

Altering conceptions of intelligence. Researchers have alleviated stereotype threat effects by altering conceptions of intelligence from static to fluid (Aronson, Fried, & Good, 2002; Good, Aronson, & Inzlicht, 2003). In one study, researchers had students consider difficulties in the seventh grade as malleable or as a result of adjustment to the seventh grade (Good et al., 2003). Female and minority students in both experimental conditions subsequently scored better on academic tests than did participants in the control condition.

This manipulation provided students with an excuse for performance other than attributing mistakes on difficult tests to innate disabilities.

Utilizing role models. Researchers have found that providing people with role models from their group helps to alleviate the effects of stereotype threat (Marx & Roman, 2002; McIntyre et al., 2003; McIntyre, Lord, Gresky, Ten Eyck, Frye, & Bond, 2005). In one study, (McIntyre et al., 2003) researchers provided stereotype-threatened women either zero, one, two, three, or four biographies to read before completing a difficult math test. Each biography described the success of one of four women in the fields or architecture, medicine, law, and invention. The researchers found that women's math performance improved as a power function of the number of biographies read, such that each successive biography improved performance by less than the previous. Women who did not read biographies of successful women scored worse than men, while women who read zero to four biographies exhibited improved performance with each additional biography. Of course, it should be noted that the women featured in the essays were successful in math-related fields. In a subsequent study (McIntyre et al., 2005), women read essays about other women who were successful in fields such as dance and art. Even when the fields of success were unrelated to mathematics, stereotype threatened women experienced improved performance while women who did not read the essays experienced impaired performance.

Providing excuses for failure. One of the reasons that stereotype threat harms performance is that the performer interprets his or her frustration on the task as indicative of the truth of the stereotype (Steele, 1997). But, when the performer has another salient explanation for that frustration, stereotype threat effects can be eliminated. Researchers have utilized a technique that allows participants an excuse for failure. For instance, one research team programmed a computer to ‘crash’ during a planned practice session, providing participants with an excuse, other than their own lack of ability, for performing poorly on a math test (Brown & Josephs, 1999, Studies 2 and 3). When stereotype threatened women were not able to practice for an up-coming math test, their performance was better than stereotype threatened women who did have the opportunity to practice for the math test.

Self-schema complexity. When people define themselves, some people have few descriptions. For instance, a man might describe himself as a husband, a father, and an employee. Other people use long descriptions including several different conceptions of self. For instance, a woman might define herself as a Christian, a student, a woman, a craft-maker, a cook, a mother, a wife, an exerciser, a worrier, and so on. Because of all of her descriptions, she could be described as someone who has a more complex self-schema than the man in the first example. Research on self-schema complexity maintains that people who have more complex self-schemas have a cognitive buffer against the negative effects of stress (Linville, 1987). People with simple self-schemas tend to have more severe self-appraisals and more severe affective responses to stressful events than people with complex self-schemas. Taking a difficult test is an example of a situation that causes stress for many people. It stands to reason that if a woman is stereotype threatened when she takes a math test, then reminding her of some other identities might help to alleviate stereotype threat effects. In one study, researchers had women consider relatively few or many of their personal identities (Gresky et al., 2005). When the women took a math test, those that considered many of their personal identities performed better than those that had considered few.

Acute and Chronic Responses to Stereotype Threat

Researchers have successfully identified a number of techniques that can be used to alleviate the effects of stereotype threat. Many members of stereotyped groups, however, have already developed strategies that may reduce the threat of poor performance. One of the responses, disengagement, is an acute response, taking place only during a particularly threatening test or situation. Repeated instances of disengagement can lead to a more chronic response, disidentification. Both of these, although helpful at alleviating threat, may lead to the perpetuation of stereotypes about group members as well as further support of the intellectual differences between groups.

Disengagement. Our self-esteem is important (Schmader, Major, & Gramzow, 2001). In order to maintain a positive evaluation of ourselves, we must either perform well in an important task, or consider that task less important. But, people who fall victim to stereotype threat are those that find the domain important (Smith & White, 2001). If these people then perform poorly on a performance measure, a likely response would be to claim that the test was not important. This pulling away from a task is known as psychological disengagement. Specifically, psychological disengagement is an acute defensive response that allows a persons' self-esteem to remain intact despite evidence of poor performance in an important domain.

To clarify, one process involved in disengagement is devaluing the domain, or considering a domain less relevant or important to the self-definition (Schmader, Major & Gramzow, 2001). In one study, women were provided with false failure feedback about their performance on a pretest of either aesthetic judgment or social sensitivity (Tesser & Campbell, 1980). After the feedback, women described the task as less related to their identities than they had indicated in an earlier questionnaire. They were also less likely to choose to complete an additional test in the area in which they failed. In other words, if they received failure feedback about the esthetic judgment task, they were more likely to choose the social sensitivity task for the second test, and if they received failure feedback about the social sensitivity task, they were more likely to choose the esthetic judgment task for the second test. These women were reducing the importance of the task by considering it less self-relevant. When people disengage often, they may move from an acute, momentary response to a situation, to a more permanent attitude toward a task: disidentification (Crocker & Major, 1989)

Disidentification. Being a member of some social groups can have important implications; for instance, within some groups, negative stereotypes are readily applied by the general public (see Crocker & Major, 1989). Despite being the target of stereotypes, researchers have found that the self-esteem of group members remains intact. One reason that this occurs is because stereotyped group members can disidentify with the stereotyped domain. For example, African Americans may consider verbal skills assessment as less important to their self-definition because the stereotype indicates that African Americans are deficient at verbal skills tasks. Women may consider mathematics less important to their self-definition because the stereotype indicates that women do not excel in math or math-related fields. Disidentification is different from disengagement in that it is a lasting conceptualization of a task or performance area. If an African American student repeatedly underperforms on a verbal skills test as a result of stereotype threat effects, he may make efforts to devalue the importance of verbal skills tests in the maintenance of his self-concept. That way, further poor performance will not harm his self-esteem. Although disidentification allows the self-esteem to remain intact, it is related to a large drop out rate among Blacks and Hispanics in high school (Griffin, 2002).

The Role of Context in Stereotype Threat Effects

Steele and colleagues (2002) mention a "theory of context" and the role that the setting may play in eliciting stereotype threat effects. They refer to the vigilance that individuals may pay to cues in a performance setting, such as the status of their identity in relation to the others present, and the likelihood that their performance will be evaluated by others. Additional stereotype threat research also highlights the role that the environment may play in eliciting stereotype threat effects (Sekaquaptewa & Thompson, 2003). In one study, researchers successfully elicited stereotype threat by displaying a poster depicting a female student struggling to complete an easy math problem while a male student easily completed a more difficult math problem (Oswald & Harvey, 2000).

Social facilitation and impairment. Social facilitation has been described as occurring when the presence of others improves performance on simple tasks (Zajonc, 1965). The underlying mechanisms of social facilitation have been revised from an explanation that includes increased "drive" and arousal, to one that suggests a self-presentational standpoint. Researchers have found that social facilitation was likely to occur when individuals completed a task that implied their ability (Bond, 1982). For instance, a woman who had stated that she was good at math and enjoyed math might experience social facilitation when asked to complete a math test in front of a crowd, especially if it became clear that she was answering many of the items correctly. On simple tasks, the woman is unlikely to make mistakes and thus her performance is consistent with the self-image she intends to portray.

She will likely receive positive feedback from onlookers and make only minor adjustments to her behavior in order to retain the approval of others.

In one study of social facilitation, Bond (1982) examined the context effects of public performance. Participants were assigned to one of two conditions. In both conditions, participants were shown consonant-vowel-consonant (CVC) trigrams and then attempted to correctly name the two syllable word that had previously been associated with the CVC in training. Participants in the *simple* condition received lists were made up of word-CVC pairings that were identical at each presentation. Thus, their task involved a predominantly easy task embedded with a few difficult items. In the *complex* condition, participants received lists that were made up of word-CVC pairings that were interchanged at each presentation, making the task predominantly difficult embedded with a few easy items. For half of the participants in each condition, an observer sat just 4 feet away and paid "polite attention" as the participants completed the task.

Bond (1982) found that simple items embedded in a difficult task are learned more slowly in the presence of an observer. In addition, participants were not affected by the presence of an observer when learning difficult items embedded in an easy task. The explanation for these findings indicates the importance of context on task learning. Bond (1982) posited that performance was not impaired for embedded complex tasks completed in front of an observer because predominantly correct performance would not prompt participants to make the strategic adjustments thought responsible for performance deficits.

Accordingly, the self-presentational view of social facilitation contends that the difficulty of the task does not directly lead to aids in performance, but rather the impression that the performer makes when completing the task (Bond, 1982). For instance, in a relevant study, researchers found that simply telling students that a test was diagnostic of ability made salient the negative stereotypes about minority groups and boosted the performance of non-minorities (Walton & Cohen, 2003). In that situation, the good impression that the majority students would make with success on the task may have come to mind and facilitated performance.

In contrast, social impairment might be seen as what happens when the context turns disadvantageous. Social impairment may occur each time a person completes a task that implies his or her inability or failure (Bond, 1982). When an individual interprets his or her public performance as failure, performance may be disrupted. For instance, in one study, researchers had participants complete a word association test in front of an observer who could see the participants, hear the participant, see and hear, or neither (Bond, Atoum, & VanLeeuwen, 1996). They found that responding to the word pairings in front of another person who could see and hear performance effectively decreased verbal learning. The researchers noted that it is not simply the presence of an observer that harms performance, but the implication that the observer has the ability to evaluate performance.

Maintaining a public image that is consistent with a desirable self-image is known as saving ‘face’ (Goffman, 1959). When a person infers that he or she is not maintaining consistent public- and self-images, the person may feel a sense of embarrassment. Embarrassment is “an indiscriminant incapacitator of continued role performance, an impediment to cognitive and motor control. Particularly disrupted will be those behaviors whose misexecution originally caused the embarrassment” (Bond, 1982, p. 1043). It follows then that social impairment is more likely to occur on difficult tasks because it is these tasks that are most likely to cause a greater number of mistakes and emphasize the evaluative nature of an observer. Again, it is not the difficulty of the task that directly inhibits performance, but the evaluation that may result from excelling or being overwhelmed by the task that may lead to alterations in performance.

Keep in mind that for some people, performing poorly should not elicit social facilitation or impairment effects. For instance, a person who had explicitly stated that he was not good at math and did not enjoy math might not experience social impairment when asked to complete difficult math problems in front of a crowd, even if it became clear that he was not completing many items correctly.

Overview

Stereotype threat research has revealed much about the types of situations that may elicit stereotype threat effects, as well as those that can alleviate stereotype threat effects (see Steele, 1997 for a review). Research on social facilitation and the importance of evaluation (Bond, 1982), however, sheds light on a new way of understanding stereotype threat effects. The current experiments were aimed at viewing stereotype threat effects through the lens of social facilitation. According to literature on social facilitation, one step toward altering performance includes inferring the evaluation that one receives (Bond, 1982). The adaptive format of the current administration of the Graduate Record Examination (GRE) provides a situation in which inferring one's progress may be easier to infer. The computer-based test provides questions based upon the performance on the test-taker. When the test-taker is performing well, the items provided by the computer become increasingly difficult. When the test-taker is performing poorly, the items provided by the computer become increasingly easy. A person completing such a test may be able to infer his or her progress based upon item difficulty. When a woman takes a test under these conditions, such a test may not only harm her performance, but harm her chances of going to graduate school. Thus, it is necessary to examine the impact of the adaptive version of the GRE.

In addition, Bond (1982) found that when difficult items are embedded in an easy task, they may be performed successfully. The explanation for these findings lies in the fact that the performer can infer that he or she is making a positive impression on an evaluative audience. At the same time, Sanna (1992) found that difficult items may be performed poorly when the performer can infer that he or she is making a negative impression on an evaluative audience. We sought to examine these effects by giving participants mostly easy or mostly difficult tests under stereotype threatening situations. We expected that as long as stereotype threatened female participants felt as though they were making a positive impression on an evaluative audience, their performance may be facilitated.

Pretest

Rationale

Previous research has found that stereotype threat effects have been more consistent when the task is difficult (Spencer et al., 1999). It was therefore important to determine the perceived difficulty of each of the potential mathematics items to be used in the following experiments. Actual difficulty, as assessed by the number of items correct, was not an important factor and was not assessed in pretesting.

Method

Participants. One-hundred twenty four undergraduate men and women participated for course credit.

Materials and procedure. One hundred and twenty multiple-choice math items were chosen from Graduate Record Examination practice booklets and Scholastic Aptitude Test

(SAT) practice booklets and organized into four 30-item packets for pretesting (Appendix A). Twelve items (three items per packet) had been altered with a computer imaging program so that they had no correct answer choice.

Participants were greeted by a female experimenter and provided with an informed consent form (Appendix B). Next, each participant received only one of the math item packets. Participants rated the level of difficulty of each item in their booklets relative to every other item on 9-point scales from 1 (*very easy)* to 9 (*very difficult*). Participants were encouraged to consider any formulas that may be required to answer each question in order to provide a precise difficulty rating for each item. The means of their ratings provided the basis of item selection for both Experiment 1 and Experiment 2.

Experiment 1

Rationale

Previous research on social facilitation highlights the effects of inferred evaluation on performance (Bond, 1982). Although people may not know exactly how others will view their performance, they may make inferences in an effort to strategically alter their behavior to maintain face. This behavioral alteration may result in improved performance on some tasks. For instance, a woman who had explicitly stated that she was good at math and enjoyed math might experience social facilitation when asked to complete an easy-for-her math test in front of a crowd, especially if it became clear to her that she was answering many of the items correctly.

The GRE is a difficult standardized test administered to college graduates, or soon-to-be graduates interested in continued education (Educational Testing Service, 2005). The GRE uses an adaptive test format that adjusts question difficulty depending upon the performance of the test-taker. If a test-taker responds correctly to a series of items, the computer will randomly choose an item that is typically more difficult than the preceding items. If a test-taker responds incorrectly to a series of items, the computer program randomly chooses an item with a rating that is slightly less difficult than the previous. To the extent that the test-taker is aware of the changing level of difficulty, the adaptive test format provides the test-taker with more feedback concerning his or her performance than a pencil-and-paper test. If a woman is performing poorly on a test of this format, an obviously easy question may provide implicit evidence that she is performing poorly without ever having received explicit feedback. Furthermore, a woman in this situation has the burden of contending with stereotype threat and confirming the negative stereotype about women's math performance, ultimately leading to poorer performance (Steele, 1997).

Experiment 1 was designed to mimic the current format of the GRE to examine whether its format may inadvertently exacerbate stereotype threat effects. In one condition, women learned that the presentation of relatively easy questions indicated poor performance, while other women learned that relatively easy questions indicated good performance. Accordingly, the hypothesis was that that women who perceived test items as an indication of poor performance would perform worse on the same test than would women who perceived easy questions as an indication of successful performance. In addition, the previous assertion includes a few underlying assumptions. For instance, the woman must consider performance in mathematics an important part of her self-concept (Smith & White, 2001). It stands to reason that a woman should not be affected by negative stereotypes about her performance if her performance hardly matters to her. She should also evaluate math item difficulty similarly to the levels intended by the researcher. That is, in order for a woman to consider a relatively easy question as an indication that she is performing well, she must first view that item as relatively easy.

Method

Participants. Eighty-two women and 21 men participated for course credit.

Materials and procedure. In order to be included in the subject pool, all participants completed an online registration process using the ExperimenTrak Research Participation System. The registration process included only the math portion of the Domain Identification Measure (Smith & White, 2001, Appendix C). High scores on the math-related items from this measure indicate that participants find mathematics important to them. Recall that in order for stereotype threat to occur, individuals must be working in a domain that they find important. All of the math identification scores were calculated and only participants who scored above the median were recruited to participate via email ($M = 24.65$, $SD = 6.64$, $Median = 25$). Of the 1,142 total staff, faculty and students who completed the registration processes, approximately 500 students were considered eligible to participate and were contacted.

During the experimental sessions, which were completed on computers using MediaLab research software (Jarvis, 2004), participants were greeted by a female experimenter. She instructed them to be seated at one of seven partitioned computer terminals. Once seated, they read and completed an informed consent form (Appendix D).

The program began with an automated "Welcome" message that appeared on the screen (Appendix E). Participants then read a set of instructions for completing a six-item practice math test (Appendix F), after which they completed the test at their own pace (Appendix G). After completing the practice test, another automated message informed participants that they were ready to take the ‘real test’ (Appendix H). The computer screen that was formerly blue changed to a black screen that displayed a large GRE logo to further distinguish the two math tests (Appendix I). Participants then read instructions designed to elicit stereotype threat. The instructions do not explicitly state that women perform better than men, but have been shown to lead participants to make that conclusion on their own (Spencer et al., 1999, Study 2).

> We are helping the Educational Testing Service to develop new items for the quantitative portion of the Graduate Record Examination (GRE-Q). As you may know, there has been some controversy about possible gender differences in math ability. The test you are about to take contains items of the type that **have** shown gender differences in the past (bold included in instructions).

Participants then indicated their gender.

The next series of screens included the instructions for the GRE. Participants read instructions about the new computer-based test which allowed for the adapting test format of the GRE. In the *punishment condition*, which was designed to mimic the current administration of the GRE, participants learned that when they responded incorrectly to several items in a row, the computer would automatically present some less difficult items that may be more appropriate for their level of ability (Appendix J). In the *reward condition,* participants were informed that when they responded correctly to several items in a row, the computer would automatically present some less difficult questions for them as a reward for correct responses. This manipulation served two purposes: First, it was important to use a test that clearly indicated to participants how they were performing during the test, rather than after the test was complete. Second, this test format design allowed administration of the same test to participants in both conditions.

Then, participants in both conditions completed a 22-item math test in 30 minutes or fewer. While completing the test, a count-down timer appeared at the bottom of the computer screen. Recall that all of the possible math items had previously been rated for difficulty, including items that had no answer (see Figure 1 for intended item layout). Ambiguous items were used at two points during the test, for items 6 and 7, and for items 15 and 16. Figure 2 displays the actual item difficulty layout based upon difficulty ratings determined in pretesting. The intent was to include items in the test for which the correct answer was unclear. That way, when the participants received a relatively easy question following the ambiguous item, they inferred, via their test instructions, how they had performed on the ambiguous items. In the reward condition, easy questions indicated that the participant had performed well on the previous, rather difficult items. In the punishment condition, on the other hand, easy questions indicated that the participant had not performed well on the previous items.

After completing the test, all participants were again shown each of the test items. Participants rated the relative difficulty of each item using a scale from 1 (*very easy*) to 9 (*very difficult*). Measuring perceived item difficulty during the experiment was important. In order for our manipulation to be effective, each of the participants needed to perceive item difficulty as they were rated in pretesting. After rating the items, students were probed for suspicion and debriefed using a standard script (Appendix K) and thanked.

Figure 1.

Intended level of Difficulty for Each Item (Experiment 1).

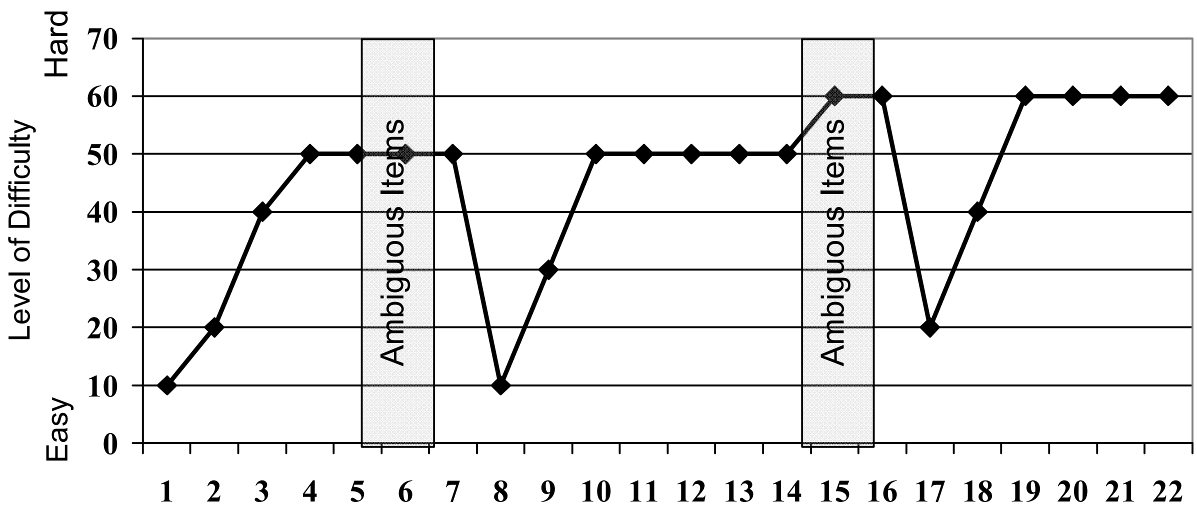
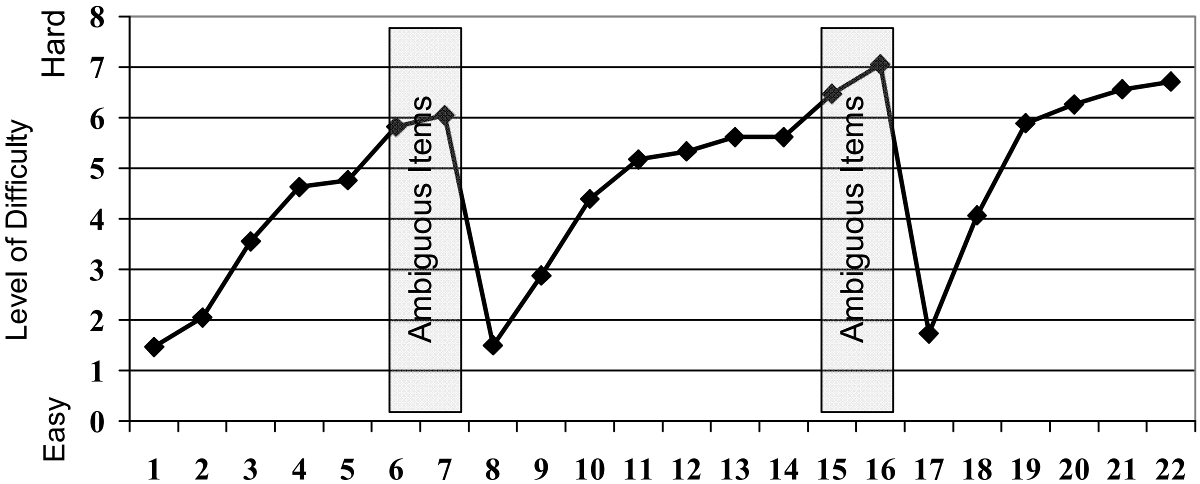


Figure 2.

Actual Difficulty Levels for Each Item as Determined During Pretesting (Experiment 1).



Results

Difficulty ratings. The central hypothesis was that when women perceived easy math items as a signal of poor performance, their math performance would suffer. In order to examine this possibility, it was necessary to first determine if participants perceived differences in item difficulty. First, a difficulty score was calculated for each of the ambiguous items (items 6, 7, 15, & 16) and a difficulty score for the two relatively easy manipulation items that followed the ambiguous items (items 8, 9, 17, & 18). Difficulty scores were analyzed with a 2 (gender) X 2 (condition: reward, punishment) X 2 (item type: ambiguous, manipulation) repeated measures ANOVA with item type as the repeated measure. The analysis yielded a significant main effect of item type, $F(1,99) = 386.13$, $p < .001$. Participants perceived the manipulation items as significantly less difficult ($M = 2.93$, $SD = 1.17$) than the ambiguous items ($M = 6.36$, $SD = 1.11$). There was no significant main effect of gender, $F(1,99) = 1.91$, $p = .28$; no significant effect of condition, $F(1,99) = .85$, $p = .359$; no significant effect of the interaction between gender and item type, $F(1,99) = .004$, $p = .951$; and no significant effect of the interaction between gender, item type, and condition, $F(1,99) = .532$, $p < .467$ (see Figures 3 & 4). The analysis yielded a significant effect of the interaction between condition and item type, $F(1,99) = 4.21$, $p < .05$ (Table 1 presents means collapsed across gender). To further investigate the interaction, simple simple effects tests were conducted on the difficulty ratings for each type of item. When participants received punishment directions, they rated ambiguous items as less difficult than participants given reward directions, $F(1,99) = 3.78$, $p = .054$. For the manipulation items, when participants were given punishment directions, they rated them no differently than participants given reward directions, $F(1,99) = .89$, *ns*.

Figure 3.

Average Difficulty Ratings for Each Math Item by Men and Women in Both Conditions (Experiment 1).

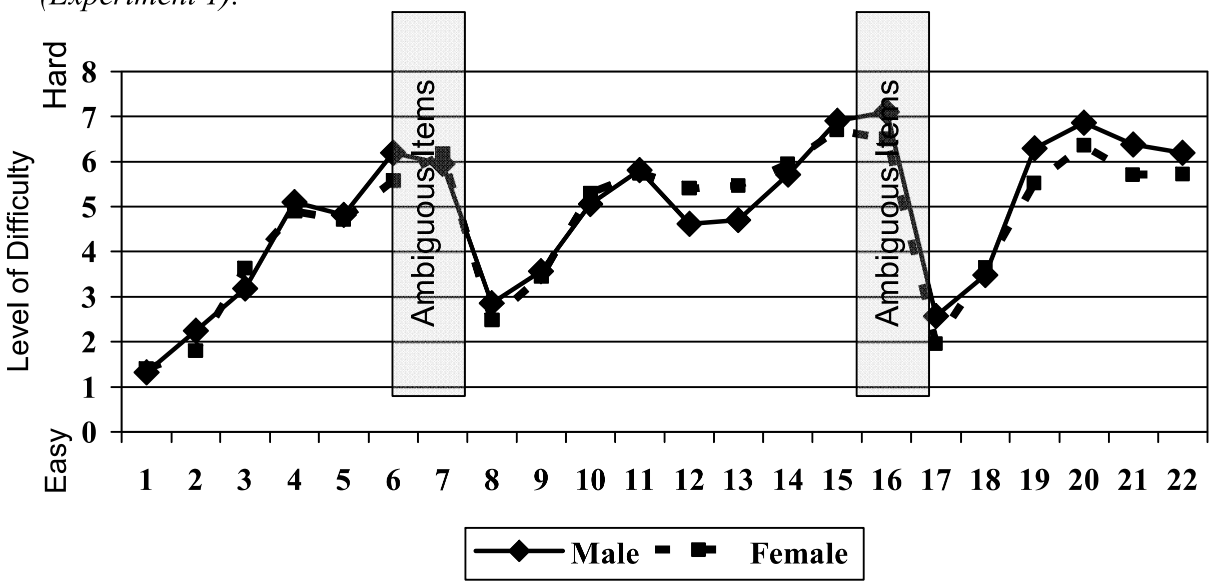


Figure 4.

Average Difficulty Ratings for Each Math Item by Women Given Punishment and Reward Instructions (Experiment 1).

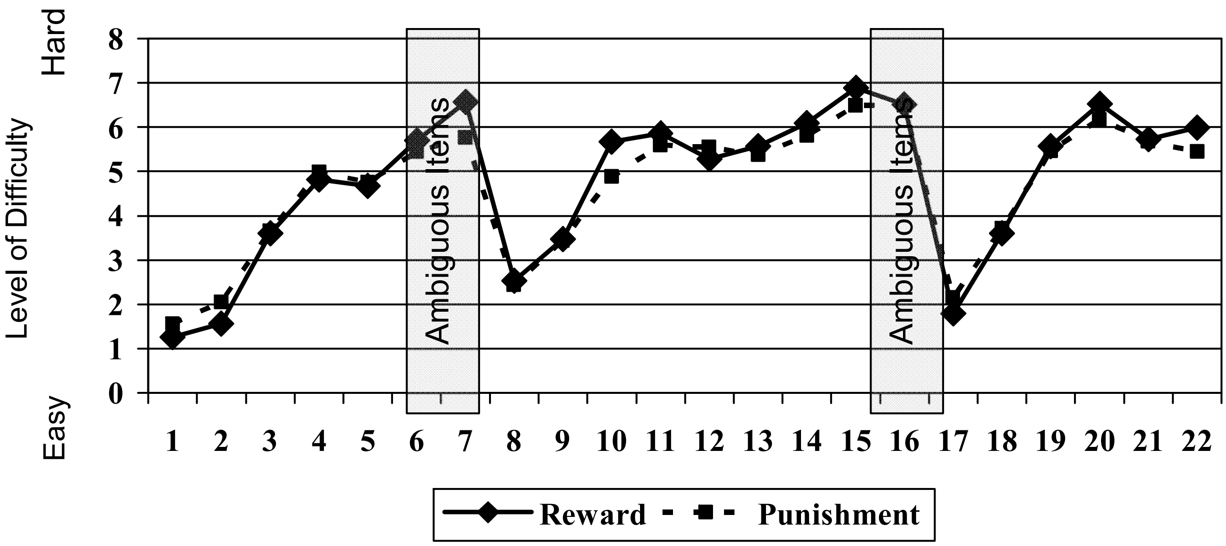


Table 1.

Mean Difficulty Ratings of Ambiguous and Manipulation Items by Participants in each Condition (Experiment 1)

Condition:	Punishment Instructions	Reward Instructions
Ambiguous	6.16[a*]	6.54[b*]
	(1.02)	(1.16)
Manipulation	3.02[c]	2.84[c]
	(1.19)	(1.15)
	N = 51	N = 52

***Note.* Standard deviations in parentheses. *p = .054.**

Women's difficulty ratings were also examined using a 2 (condition: reward, punishment) X 2 (item type: ambiguous, manipulation) repeated measures ANOVA treating item type as the repeated measure (Table 2). The ANOVA yielded a significant effect of item type, such that women rated ambiguous items as significantly more difficult that manipulation items, $F(1, 80) = 419.78$, $p < .001$. The effect of condition was not significant, $F(1,80) = .06$, p = .80. The interaction between item type and condition was not significant, $F(1,80) = 1.91$, $p = .17$.

Percent correct. Completing easy mathematics questions after completing difficult items in the punishment condition should have caused participants to infer that their overall performance was poor (see Figures 5 & 6). Recall that the hypothesis stated that, for females, indicating one's gender before such a test should elicit stereotype threat effects and they would perform more poorly on target items than women with reward directions. The percentage of correct responses for items 10-14 and 19-22 was examined in a 2 (gender) X 2 (condition: reward, punishment) multivariate analysis of variance (MANOVA) which yielded one overall main effect of gender, $F(2,98) = 16.741$, $p < .001$. Males scored significantly better on the combined total of items 10-14 and 19-22 ($M = 39.80\%$) than females ($M = 23.50\%$). The MANOVA did not yield a main effect of condition for items 10-14 or 19-22, Fs < .56, ps > .46. The MANOVA did yield a main effect of gender for items 10-14, $F(2,98) = 9.77$, $p < .01$. Men earned a higher percent correct score on items 10-14 ($M = 49.52\%$, $SD = 24.18\%$) than did women ($M = 30.73\%$, $SD = 24.78\%$). The MANOVA also yielded a main effect of gender for items 19-22. Men earned higher percent correct scores on items 19-22 ($M = 29.76\%$, $SD = 26.94$) than did women ($M = 16.16\%$, $SD = 21.65\%$; $F(2,98) = 5.86$, $p <$

Table 2.

Mean Difficulty Ratings of Ambiguous and Manipulation Items by Women in Each Condition (Experiment 1)

Condition:	Punishment Instructions	Reward Instructions
Ambiguous	6.17[a]	6.45[a]
	(1.02)	(1.16)
Manipulation	2.98[b]	2.79[b]
	(1.24)	(1.12)
	N = 40	N = 42

***Note.* Standard deviations in parentheses.**

Figure 5.

Average Percent Correct for Each Math Item by Men and Women (Experiment 1).

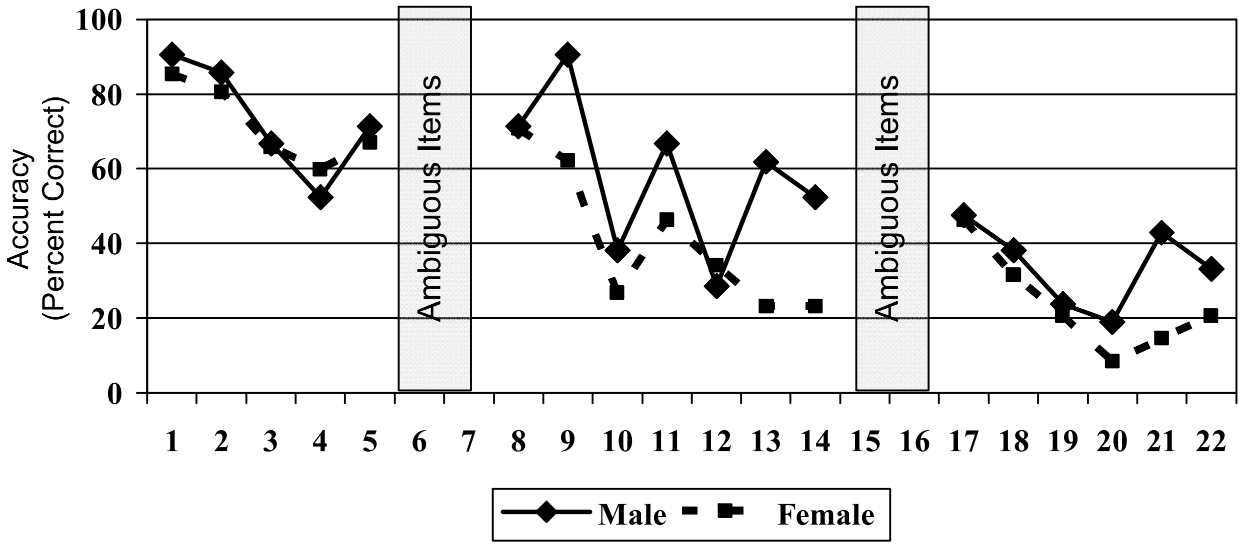
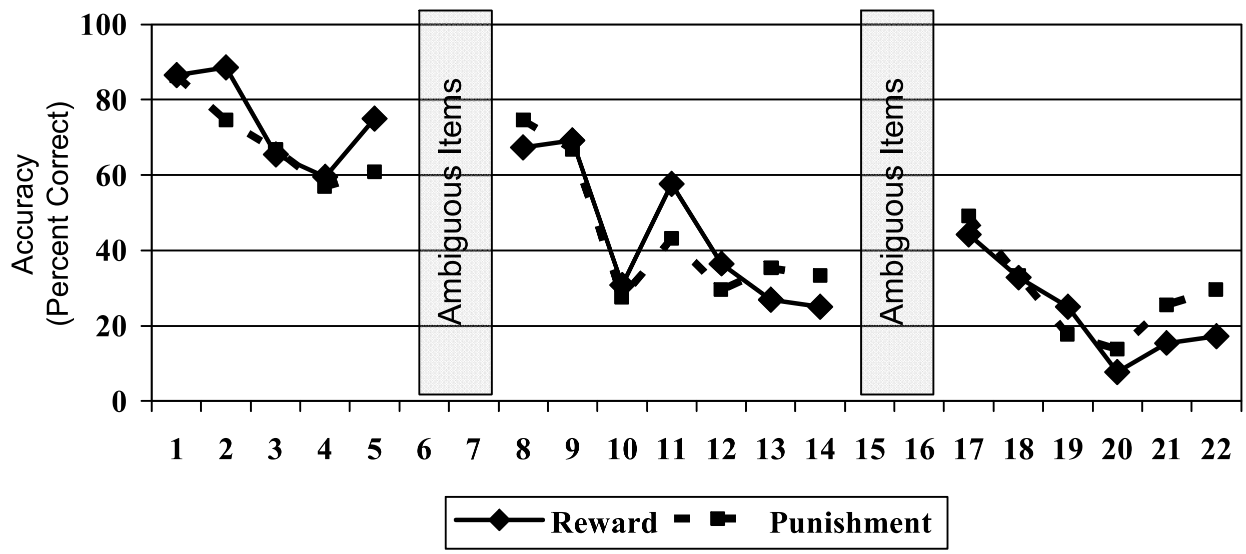


Figure 6.

Average Percent Correct for Each Math Item by Participants in Each Condition (Experiment 1).



.05). The interaction between gender and condition was non-significant for both items 10-14 and items 19-22, *F*s < .45, *p*s > .50.

Next, planned comparisons were conducted for the top two rows of means shown in Table 3. For women, there was no significant effect of testing condition on their scores for items 10-14, $t(81) = .07$, $p = .47$. Also, women's performance on items 19-22 was not affected by test directions, $t(81) = 1.21$, $p = .12$ (see Figure 7).

Research on stereotype threat often includes a method designed to equate groups for any initial differences in math ability (Aronson et al., 2002; Spencer et al., 1999, Study 3; Steele & Aronson, 1995). In the current experiment, the six-item practice test provided a measure of math ability prior to stereotype threat induction. Practice test scores were used as the covariate in a 2 (gender) X 2 (condition: reward, punishment) analysis of covariance (ANCOVA) of the first dependent variable, items 10-14. When controlling for practice test score, the main effect of gender remains significant, $F(1,98) = 10.15$, $p < .01$. Next, a 2 (gender) X 2 (condition: reward, punishment) ANCOVA of the second dependent variable, items 19-22, was conducted using percentage correct of practice test scores as the covariate. The ANCOVA main effect of gender also remained significant on these items when controlling for practice test scores, $F(1,98) = 6.31$, $p < .05$.

Another potential covariate for examining men's and women's scores on the math test is their responses to items 1-5. Because items 1-5 occur before the first manipulation, but after the stereotype threat induction, the results should be interpreted with caution. Performance on items 10-14 was examined by conducting a 2 (gender) X 2 (condition: reward, punishment) ANCOVA using the percent correct scores for items 1-5 as a covariate. The main effect of gender remained significant when controlling for scores on items 1-5 in

Table 3.

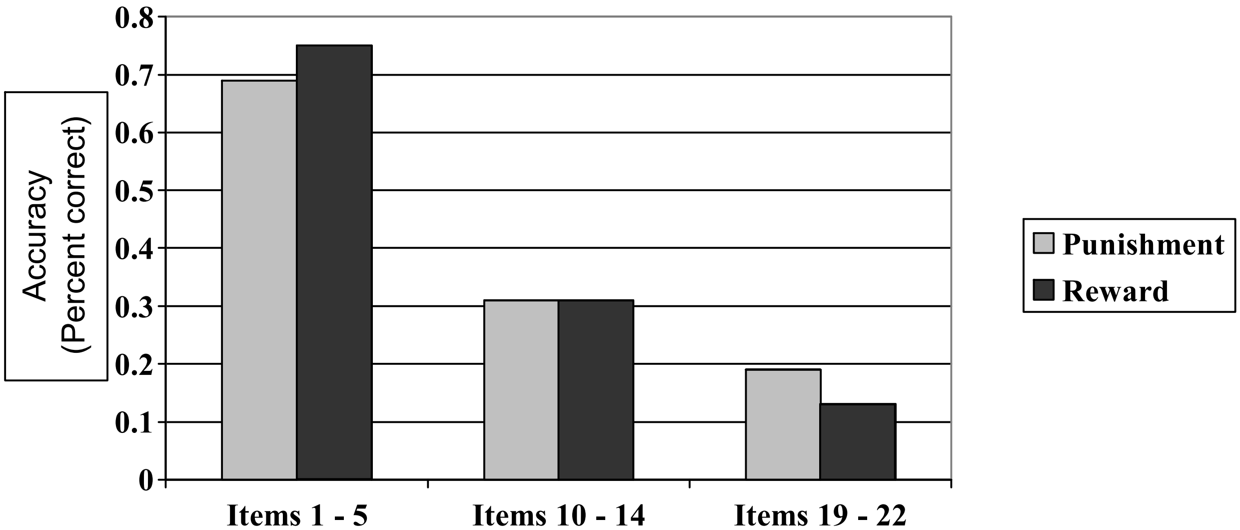
Mean Percentage Correct Responses for Each Section of Questions for Men and Women (Experiment 1)

Condition:		Punishment Instructions	Reward Instructions
Women			
	Items 10-14	30.50[a]	30.95[a]
		(26.79)	(23.03)
	Items 19-22	19.38[b]	13.10[b]
		(24.34)	(18.51)
		N = 40	N = 42
Men			
	Items 10-14	45.45[c]	54.00[c]
		(25.42)	(23.19)
	Items 19-22	29.55[d]	30.00[d]
		(21.85)	(32.91)
		N = 11	N = 10

***Note.* Standard deviations in parentheses.**

Figure 7.

Percentage of Correct Items in Each Section of the Math Test by Women (Experiment 1).



an analysis of items 10-14, $F(1, 98) = 10.42$, $p < .01$, and in an analysis of items 19-22, $F(1,98) = 5.93$, $p < .05$. The main effect of reward versus punishment condition and the interaction remained non-significant in all these ANCOVAs.

Domain identification. All of the men and women who participated in the present experiment were recruited because they reported that math was important to their sense of self (Smith & White, 1999). Linear regression was used to examine the moderating effects of domain identification on condition for percent correct scores for items 10-14 and 19-22. Before conducting the regression analysis, participant’s domain identification scores were transformed into *z*-scores (Baron & Kenny, 1986). The linear regression model included gender, condition, domain identification, and the interaction between each of those variables (gender X condition, gender X domain identification, condition X domain identification, gender X condition X domain identification) as predictors of the percent correct scores for items 10-14 (Table 4 includes inter-correlations of the predictors). The analysis revealed that gender was the only significant predictor of percent correct scores on items 10-14 (Table 5). The model significantly predicted scores on items 10-14, $F(7,102) = 2.93$, $p <$ 01. In other words, when gender, domain identification, and condition are known, one can predict scores on items 10-14.

A second regression analysis was conducted to predict scores on items 19-22 using the same predictor variables noted above. The analysis revealed no significant relationships between the predictors and performance on items 19-22 (Table 6). The model, including all possible predictors was a marginally significant predictor of scores on items 19-22, $F(7,102) = 1.81$, $p = .09$.

Table 4.

Means, Standard Deviations, and Correlations of Target Math Items and Gender, Condition, and Domain Identification (Experiment 1; N = 103) *o*

	1	2	3	4	5
PerCor10-14					
PerCor19-22	.00				
Gender	.30**	.24*			
Condition	.03	-.11	-.03		
Domain Identification	.23*	.07	-.06	-.02	
M	34.56%	18.93%	.20	.50	.00
SD	25.70%	23.34%	.41	.50	1.00

*** *p* < .05; ***p*<.01; PerCor10-14 = percent correct on items 10-14, PerCor19-22 = percent correct on items 19-22; M = mean; SD = standard deviation.**

Table 5.

Linear Regression Predicting Percent Correct Scores on Items 10-14 from Gender, Condition, Domain Identification, and Their Interactions (Experiment 1)

	B	SE	Beta (β)	t	P-value
Constant	.30	.04		7.86	.00
GEN	.19	.08	.30	2.29	.02
CON	.00	.05	.02	.17	.86
DOMID	.05	.04	.18	1.24	.22
GEN X CON	.04	.12	.05	.33	.74
GEN X DOMID	.16	.10	.25	1.63	.11
CON X DOMID	.01	.05	.02	.14	.89
GEN X CON X DOMID	-.14	.13	-.17	-1.10	.27

Note. **$F(7,102) = 2.93$, $p < .01$. GEN = gender; CON = condition; DOMID = domain identification score; SE = standard error.**

Table 6.

Linear Regression Predicting Percent Correct Scores on Items 19-22 from Gender, Condition, Domain Identification, and Their Interactions (Experiment 1)

	B	SE	Beta (β)	t	P-value
Constant	.19	.04		5.43	.00
GEN	.12	.08	.21	1.54	.13
CON	-.06	.05	-.14	-1.28	.20
DOMID	-.02	.04	-.10	-.63	.53
GEN X CON	.05	.11	.07	.46	.65
GEN X DOMID	.14	.09	.24	1.51	.14
CON X DOMID	-.04	.05	.12	.80	.43
GEN X CON X DOMID	-.04	.12	-.05	-.30	.77

***Note.* $F(7,102) = 1.82$, $p = .09$GEN = gender; CON = condition; DOMID = domain identification score; SE = standard error.**

Linear regression was also used to predict women's scores only, because it is women for whom we expected stereotype threat to occur (see Table 7 for inter-correlations of the predictors). The model included domain identification, condition, and condition X domain identification as predictors of women's scores on items 10-14. The analysis yielded no significant relationships between percent correct on items 10-14 and the predictor variables (Table 8). The model did not significantly predict women's scores on items 10-14, $F(3,81) = 1.16$, $p = .33$. A second regression was conducted to predict women's scores on items 19-22 using domain identification, condition, and condition X domain identification as predictors. The analysis yielded no significant relationships between percent correct on items 19-22 and the predictor variables (Table 9). The model did not significantly predict women's scores on items 19-22, $F(3,81) = .80$, $p = .50$.

Discussion

It was expected that the manipulation, providing success feedback in the form of relatively easy questions, would facilitate subsequent math performance for women under stereotype threat. Under these circumstances, women would interpret the easier questions as evidence of their successful performance and use that evidence as a rebuttal for the negative stereotype about their group. Then, the women would be able to perform better than women who were provided with failure feedback, and equally to men in both conditions.

One possible result of the manipulation is that women might not have attributed their success on the ambiguous questions internally. It is possible that women in the reward condition interpreted their success on the difficult ambiguous questions as luck. According to balance theory (see Gilmor & Minton, 1974), when a person's assessment of his or her ability is low, then success should be attributed to external sources such as luck or fate, and

Table 7.

Means, Standard Deviations, and Correlations of Target Math Items, Condition, and Domain Identification for Women Only (Experiment 1; N = 103) *o*

	1	2	3	4
PerCor10-14				
PerCor19-22	-.18			
Condition	.01	-.15		
Domain Identification	.21	.00	-.05	
M	30.73%	16.16%	.51	.03
SD	24.78%	21.65%	.50	1.02

PerCor10-14 = percent correct on items 10-14, PerCor19-22 = percent correct on items 19-22; M = mean; SD = standard deviation.

Table 8.

Linear Regression Predicting Percent Correct Scores on Items 10-14 from Condition, Domain Identification, and Their Interactions for Women Only (Experiment 1)

	B	SE	Beta (β)	t	P-value
Constant	.30	.04		7.70	.00
CON	.01	.06	.02	.17	.87
DOMID	.05	.04	.19	1.21	.23
CON X DOMID	.01	.05	.02	.14	.89

Note. ***F*****(3,81) = 1.16,** ***p*** **= .33; CON = condition; DOMID = domain identification score; SE = standard error.**

Table 9.

Linear Regression Predicting Percent Correct Scores on Target Items from Condition, Domain Identification, and Their Interactions (Experiment 1)

	B	SE	Beta (β)	t	P-value
Constant	.19	.03		5.67	.00
CON	-.06	.05	-.15	-1.34	.19
DOMID	-.02	.03	-.10	-.66	.51
CON X DOMID	-.04	.05	.13	.83	.41

***Note.* F(3,81) = .80, p = .50; CON = condition; DOMID = domain identification score; SE = standard error.**

failure should be attributed internally, to lack of ability or effort. It also follows that when a person's assessment of his or her ability is high, success should be attributed to internal sources, such as ability or effort, and failure should be attributed to external sources. In a related study (Gilmor & Minton, 1974), researchers introduced an anagram task to students. Before completing the task, students rated their confidence about their ability to perform well on the task. After completing the task and receiving feedback on their performance, all of the students indicated the direction of their attribution; internally or externally. For initially high confident students, news of success led them to make an internal attribution. But, for initially low confident students, news of their success led them to make an external attribution. In other words, students with low initial confidence did not claim responsibility for their success. In a second study, researchers asked students to rate how their performance was related to their ability, a factor that was not included in the previous study (Feather & Simon, 1964). They found that initially low confident people who were provided with success feedback did not consider their own ability an important factor in determining their success.

Taken together, the previously reported studies support the possibility that women in the reward condition may have interpreted their success on the difficult questions to luck rather than to their ability to complete difficult math problems. This assertion may seem to contradict the fact that the men and women who participated in our study indicated that they not only found math important, but they performed well in areas of mathematics. But, research in stereotype threat indicates that before a stereotype threatening performance, the performance expectations held by stereotyped group members are low (Stangor et al., 1998). Thus, despite the fact that they generally have high expectations concerning their math performance, women may have low expectations in a stereotype threatening situation. Their low expectations may have led them to interpret their success feedback as unrelated to their ability, but rather the result of luck. As a result, receiving easy questions would not have been perceived as a rebuttal for the negative stereotype about women and would not have consequently improved women's performance.

Overall, the intent of the experiment was to demonstrate that within test feelings of failure can lead to impaired performance during a math test for stereotype threatened women, but that within test feelings of success can lead to improved performance. Women may have felt successful in the reward condition, but not as a result of their own ability or talent. Recall that the items without a correct answer were used as the manipulation so that, in either condition, test-takers could use the following easy question to reflect on their previous response. Choosing to use the ambiguous items for the manipulation may have spoiled this process. One caveat to the approach is that test-takers may be absolutely sure that they have completed the ambiguous problem correctly but they do not see a matching answer option. A better strategy would be to present practice test problems with a "difficulty meter" on the side of the screen that has inflated (or deflated) difficulty ratings. That way, if the test-taker does determine the correct response, the correct answer will be among the answer choices.

Experiment 2

Rationale

Recall that stereotype threat effects are most often found for relatively difficult questions. For the woman completing a math test under stereotype threat, difficult items may cause her to expend extra effort to find the correct answer. When she uses extra effort, it may then lend support to the negative stereotype about her group. Research in social facilitation, however, suggests that under certain conditions, women might successfully complete difficult questions. Bond (1982) found that participants were able to learn difficult items in the presence of an observer, if those difficult items were embedded in an easy task. The opposite was also true: participants who learned simple items embedded in a difficult task learned the simple items more slowly in the presence of an observer. The explanation for these findings supports the effect of context on task learning. Difficult items may be perceived as less difficult when they are presented in a relatively easy context. The current experiment employed a similar design, in which men and women completed difficult items that were embedded in either a predominantly easy or a predominantly difficult task. The hypothesis was that as stereotype threatened women came across a difficult question in an easy task, their performance should not suffer the typical effects associated with stereotype threat. It was also hypothesized that as stereotype threatened women encountered the same difficult questions embedded in a predominantly difficult task, their performance should suffer as a result of stereotype threat.

Participants. Seventy eight women and 22 men participated for course credit.

Materials and procedure. Participants in Experiment 2 were chosen using the same process used in Experiment 1. After participants completed the online registration process using the ExperimenTrak Research Participation System, their scores on the Domain Identification measure were collected (Smith & White, 2001, Appendix C). The recruited participants had mathematics identification scores in the top half of the distribution.

During the experimental sessions, participants were seated at one of seven partitioned computer terminals by a female experimenter. Once seated, they read and completed an informed consent form (Appendix D).

Each of the participants completed a computer program using MediaLab research software (Jarvis, 2004). The program began with an automated "Welcome" message that was identical to the message used in Experiment 1 (Appendix E). Participants then read a set of instructions for completing a six-item practice math test (Appendix F), after which they completed the test at their own pace (Appendix L). After completing the practice test, another automated message informed participants that they were ready to take the 'real test' (Appendix H). The computer screen that was formerly blue changed to a black screen that displayed a large GRE logo to further distinguish the two math tests (Appendix I). Participants then read a set of standard stereotype threat instructions:

> We are helping the Educational Testing Service to develop new items for the quantitative portion of the Graduate Record Examination (GRE-Q). As you may know, there has been some controversy about possible gender differences in math ability. The test you are about to take contains items of the type that **have** (bold included in instructions) shown gender differences in the past.

Participants then indicated their gender.

The next series of screens included the instructions for the GRE. Participants read instructions about the change from pencil-and-paper based tests to computer-based tests and the "diagnosticity" of the GRE. In the *difficult in difficult condition*, participants completed a 22-item math test in 30 minutes or fewer (Appendix M). While they completed the test, a count-down timer appeared at the bottom of the screen.

All of the items selected for use in Experiment 2 were pretested for perceived difficulty on a scale that ranged from 1 (*very easy*) to 9 (*very difficult*). All of the items used in the difficult in difficult condition had been rated as moderate to difficult ($M = 4.63$ or higher).

In the difficult in easy condition, participants had up to 30 minutes to complete a math test that included 17 relatively easy questions (M_{range} 1.90-3.68) as well as five difficult items. Items 6, 11, 15, 19, and 22 were identical to the items in the difficult in difficult condition and were located in the same serial position (Appendix M).

After completing the test, participants were probed for suspicion and debriefed using a standard script (Appendix K), and thanked.

Results

Because of a computer programming error, some of the data for Experiment 2 were lost. Other researchers have experienced a similar problem with this program (Tormala, Clarkson, & Petty, 2006). The first step in examining data involved calculating three separate totals: the total correct, the total incorrect, and the total impossible answers[1]. Table 10 displays the means for each of the target items. In the first column, Item 6 includes a relatively large number of incorrect responses and impossible answers. Also, unlike the other target items, Item 6 was negatively correlated with the total number of correct responses across all five target items ($r = -.30$, $p < .05$). For this reason, item 6 is not included in the following analyses.

Manipulation check. In the difficult in difficult condition, it was expected that participants would earn lower scores on the non-target items than participant's scores in the difficult in easy condition (Table 11). A 2 (gender) X 2 (condition: difficult in difficult, difficult in easy) ANOVA of the total non-target items correct yielded a significant main effect of condition, $F(1,96) = 134.05$, $p < .001$. Participants completing mostly easy

Table 10.

Total Number of Correct, Incorrect, and Impossible Answers on Each of the Shared Math Items (Experiment 2)

Item:	6	11	15	19	22
Correct	8	63	52	40	24
Incorrect	58	37	46	55	61
Impossible	34	0	2	5	15
	N = 100	N = 100	N = 100	N = 100	N = 100

Table 11.

Mean Number of Non-Target Math Items Correct by Women and Men (Experiment 2)

Condition:	Difficult in Difficult	Difficult in Easy
Women	3.97[a]	9.47[b]
	(1.96)	(1.93)
	N = 36	N = 42
Men	4.44[c]	11.00[d]
	(2.28)	(1.09)
	N = 16	N = 6

***Note.* Standard deviations in parentheses. Total possible items = 17.**

questions responded correctly more often ($M = 9.67$, $SD = 1.91$) than participants completing mostly difficult questions ($M = 4.12$, $SD = 2.05$).

Number correct. Next, scoring on the four target items was examined by conducting a 2 (gender) x 2 (condition: difficult in difficult, difficult in easy) ANOVA, which yielded a significant main effect of condition, $F(1,96) = 6.55$, $p < .05$. The effect of gender was non-significant, $F(1,96) = .77$, $p = .39$. The gender X condition interaction was also non-significant, $F(1,96) = .07$, $p = .79$.

Next, simple effects tests were conducted for both men's and women's scores on the target items (Table 12). Women performed significantly better on the target items when the context questions were easy rather than difficult, $F(1,96) = 7.17$, $p < .01$. Men did not perform better when the context questions were easy rather than difficult, $F(1,96) = 2.45$, *ns*. Women in the difficult in easy condition selected more correct responses on the four target items than did women in the difficult in difficult condition. Although research on women's mathematics stereotype threat generally examines number correct and percentage of attempted items correct as a measure of performance (e.g., Ambady et al., 2004; Davies et al., 2002; Inzlicht & Ben-Zeev, 2003; Keller, 2002; Schmader, 2002; Shih et al., 1999), examining both measures of performance in the current experiment would have been redundant because all of the participants completed the math test.

Response time. It is possible that participants in the difficult in easy condition spent more time on the target items than did participants in the difficult in difficult condition. An average time for responding to each of the target items and each of the non-target items was computed. Using the averaged response times as a repeated measure, a 2 (gender) X 2 (condition: difficult in difficult, difficult in easy) X 2 (item type: target, non-target) repeated

Table 12.

Mean Number Items Correct on the Four Shared Math Items by Women and Men (Experiment 2)

Condition:	Difficult in Difficult	Difficult in Easy
Women	1.47[a]	2.05[b]
	(.85)	(.93)
	N = 36	N = 42
Men	1.63[c]	2.33[c]
	(1.08)	(1.21)
	N = 16	N = 6

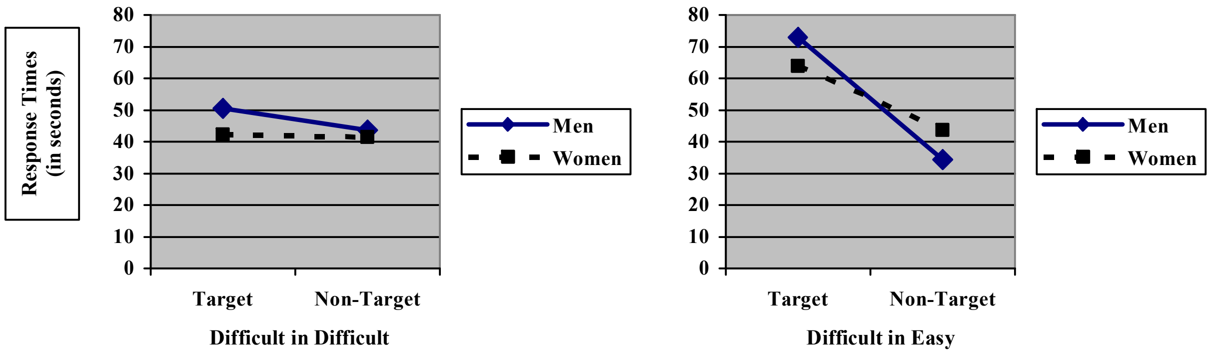
***Note.* Standard deviations in parentheses. Shared math items include items 11, 15, 19, and 22.**

measures ANOVA was performed. The ANOVA yielded a main effect of item type (see Figure 8). Participants took significantly longer to respond to the target items than to the non-target items, $F(1,94) = 27.03$, $p < .001$. The ANOVA yielded a marginally significant interaction between item type and gender, $F(1,94) = 3.69$, $p = .06$. The ANOVA also yielded a significant interaction between item type and condition, $F(1,94) = 16.09$, $p <.001$. The three-way interaction was non-significant, $F < .93$. When simple effects tests were performed, the interactions became more clear. In the difficult condition, there were no significant differences between men's and women's response times to either type of item, *F*s < 1.60, *p*s > .21. In the easy condition, there was no significant difference between men's and women's response times to the target items, $F(1,47) = .60$, $p = .44$. Response times for the non-target items were different for men and women in the easy condition. Men took significantly less time to respond to the non-target items than women in this condition, $F(1,47) = 4.47$, $p < .05$.

Next, women's average response times for the target and non-target items were examined with a 2 (condition: difficult in difficult, difficult in easy) X 2 (item type: target, non-target) repeated measures ANOVA with item type as the repeated measure, which yielded a significant main effect of item type. Women took significantly longer to respond to target than to non-target items, $F(1,76) = 13.69$, $p < .001$. The ANOVA also yielded a significant interaction between item type and condition, $F(1,76) = 11.84$, $p = .001$. Simple effects tests were performed to examine women's response times further. The ANOVA indicated that women in the difficult in easy condition took significantly longer to respond to target items ($M = 63.90$) than women in the difficult in difficult condition ($M = 42.17$; $F(1,77) = 14.77$, $p < .001$). Women in the difficult in easy condition did not, however take

Figure 8.

Mean Response Times in Seconds on Target and Non-target Items for Men and Women in Each Condition (Experiment 2).



significantly longer to respond to non-target items (*M* = 43.72) than women in the difficult in difficult condition (*M* = 41.44; *p* = .39). The correlation between women's average response times on the target items and their percent correct scores on the same items was also examined. The Pearson product moment correlation revealed a significant positive relationship between the two variables, r = .44, $p < .001$. The correlation between non-target item response times and percent correct on non-target items was not significant, r = .02, *p* = .82.

Women's percent correct scores on the target items in each condition were examined with an ANCOVA, using time to complete each item as the covariate. The results of the analysis reveal that main effect of condition is no longer significant when we account for the amount of time women spent completing target items, $F(1,77) = 1.927$, *p* = .169. Although women in the easy condition took longer to respond to target items than women in the difficult condition, women in the difficult condition took longer to complete the total test.

Domain identification. As with Experiment 1, the effect of domain identification on the target items was also examined. Linear regression was used to examine the moderating effects of domain identification on target items 11, 15, 19, and 22. The model included gender, z-scores for domain identification, condition and the interaction between each of those variables (gender X domain identification, gender X condition, domain identification X condition, gender X domain identification X condition) as predictors (see Table 13 for inter-correlations of predictors). Condition was significantly related to performance on the target items (Table 14). All other predictors were not significantly related to percent correct scores. A second linear regression analysis included women's scores only. The model included domain identification, condition, and domain X condition as predictors of women's

Table 13.

Means, Standard Deviations, and Correlations of Target Math Items and Gender, Condition, and Domain Identification (Experiment 2; N = 100)

	1	2	3	4
PerCorTarget				
Gender	.02			
Condition	.29**	-.22*		
Domain Identification	.024	.05	.15	
M	45.75%	.22	.48	.00
SD	24.43%	.42	.50	1.00

*** *p* < .05; ***p*<.01; PerCorTarget = percent correct on target items; M = mean; SD = standard deviation.**

Table 14.

Linear Regression Predicting Percent Correct Scores on Target Items from Gender, Condition, Domain Identification, and Their Interactions (Experiment 2)

	B	SE	Beta (β)	t	P-value
Constant	.37	.04		8.94	.00
GEN	.03	.07	.06	.46	.65
CON	.15	.06	.31	2.63	.01
DOMID	-.01	.04	-.04	-.22	.83
GEN X CON	.02	.13	.02	.11	.90
GEN X DOMID	.04	.07	.08	.52	.61
CON X DOMID	-.02	.06	-.06	-.37	.71
GEN X CON X DOMID	.03	.11	.03	.23	.82

Note. **F(7,94) = 1.43, p = .20. GEN = gender; CON = condition; DOMID = domain identification score; SE = standard error.**

performance on the target items (see Table 15 for inter-correlations of predictors). Of the predictors, only condition significantly predicted percent correct scores for women (Table 16).

Discussion

Consistent with the self-presentational view of social facilitation (Bond, 1982), women were able to perform better on difficult items embedded in an easy task. When any concerns arose about the impression that they could be making on the experimenter, they could reflect on their performance throughout the test. For women in the other condition, however, their public- and self-images were in conflict. They had already reported that mathematics was important. Then, when they completed math problems that were threatening to their self-image, they may have experienced a sense of embarrassment that inhibited performance.

Addition social facilitation research supports the current findings and the emphasis on the possibility of evaluation. In the current study, participants were told that their results on the test were “indicative of their ability” in an effort to increase their awareness of the evaluative nature of the test. In one study, participants were given either a relatively difficult version or relatively easy version of the Remote Associates Test (RAT; Sanna, 1992). After completing the list, participants completed a number of questions. The results indicated that the more difficult test yielded lower performance expectations for the participants, especially when they believed that their performance was being evaluated. In addition, when participants believed that they would be evaluated, their performance on the difficult test was significantly worse than participants who believed that their performance could not be

Table 15.

Means, Standard Deviations, and Correlations of Target Math Items, Condition, and Domain Identification for Women Only (Experiment 2; N = 78)

	1	2	3
PerCorTarget			
Condition	.31**		
Domain Identification	-.03	.16	
M	44.55%	.54	-.03
SD	23.37%	.50	.97

*** *p* < .05; ***p*<.01; PerCorTarget = percent correct on target items; M = mean; SD = standard deviation.**

Table 16.

Linear Regression Predicting Percent Correct Scores on Target Items from Condition, Domain Identification, and Their Interactions (Experiment 2)

	B	SE	Beta (β)	t	P-value
Constant	.37	.04		9.60	.00
CON	.15	.05	.33	2.83	.00
DOMID	-.01	.04	-.04	-.22	.81
CON X DOMID	-.02	.06	-.06	-.40	.69

Note. **$F(3,72) = 2.77$, p=.05. CON = condition; DOMID = domain identification score; SE = standard error.**

evaluated by the experimenter. On the other hand, the possibility of evaluation boosted scores of the participants.

Another interpretation of our findings is that women completing the difficult items embedded in a difficult task performed worse because of the amount of time they spent working on the target math items. In other words, they may have performed worse because they were unable to expend enough effort. One line of research in stereotype threat presents evidence that withdrawal of effort does not play a significant role in accounting for stereotype threat effects on performance (McKown & Weinstein, 2003). To be sure that withdrawal of effort does not play a role in future research, only math items that have been extensively pretested will be use. The pretested questions included in the predominantly easy and a predominantly difficult test should require the same amount of time to complete, without regard for the actual number of items on the test.

Currently, stereotype threat alleviation involves techniques such as describing a test as non-diagnostic (Steele & Aronson, 1995); inducing individuation (Ambady et al., 2004); providing excuses for failure (Brown & Josephs, 1999); and making salient many identities (Gresky et al., 2005). Unfortunately, the most likely goal of stereotype threat researchers is to discover ways to apply alleviation techniques in the classroom to equalize scores between minority and majority groups. All of the previously mentioned techniques, though effective, would be difficult to implement in a classroom setting. The current experiment brings us one step closer to a classroom-friendly technique. Math teachers could use math tests that contain a combination of relatively easy and relatively difficult questions when assessing their students' ability. In this manner, all students would be free to provide their best performance.

GENERAL DISCUSSION

Academic performance is important. The current experiments were intended to examine two in-test techniques to alleviate stereotype threat effects for women (see Steele, 1997). In the first experiment, it was proposed that perceiving one's performance as successful early in a task might improve performance on that task for women under stereotype threat. Although the hypothesis was unsupported with the current methodology, literature on social facilitation (Bond, 1982) indicates that, with changes, finding support for the hypothesis is still plausible. In other words, creating a situation in which women feel successful at math, as a result of genuine ability, rather than luck, should allow them to save 'face' and exhibit better performance on a math test. Future research is required to determine which methods would most effectively achieve this outcome.

In the second experiment, the hypothesis stated that women would perform better on difficult math questions that are embedded in an easy test than women completing the same question embedded in a more difficult test. The hypothesis was supported. It was expected that women when women brought to mind the stereotypical expectations for their performance, they could also bring to mind their performance on the easier items on the test. In doing so, women would be able to provide concrete evidence that they have been able to save 'face' and avoid a conflict between their public- and self-images. Thus, their performance on the math test would be unaffected by the threat of stereotype applicability. Future refinements of the current methodology are required to clarify findings, because the easy context also helped men to perform better.

Stereotype Threat in the Classroom

In a recent decision, the United States Department of Education will now allow single-sexed schools and classrooms, assuming that they are equal in quality. While a possible drawback to this decision includes the possibility of increasing discrimination and prejudice regarding sex differences, it is possible that in those classrooms, stereotype threat may be overcome. This assertion comes from two sources. First, in one study, I attempted to elicit stereotype threat effects among African American students at a predominantly (73%) Black college, but I was unsuccessful. I reasoned that in predominantly black colleges, their peers act to reduce the applicability of stereotypes and alleviate stereotype threat effects during many test situations (see Cadinu et al., 2003). Other researchers have also found that stereotype threat was not operative in predominantly black colleges (Smith & Hopkins, 2004). They reasoned that, despite knowledge of the stereotypes, the intellectual environment may have reduced the possible veracity of social stereotypes for African Americans. The same possibility exists for young girls and teens entering into exclusively female educational institutions. In such a learning setting, girls and teens may have daily examples that reduce the applicability of negative stereotypes concerning their performance. It is possible that stereotypes in our society are so pervasive that even women in same-sex schools can not out run their reputation, but future research will surely test whether having a same-sex social context facilitates or impairs performance, just as the present research tested whether having an easy item context facilitates performance.

Future Directions

African American high school students. In previous research, it was found that women's mathematics stereotype threat could be alleviated by reminding them of other women who have been successful within and without the domain of mathematics (McIntyre et al., 2005). This alleviation technique is intriguing because it may be uniquely successful for women. African American males tend to chose role models outside of education (Assibey-Mensah, 1997). In fact, in a study of 4,500 African American boys between the age of 10 and 18 years of age, between 85 and 95% of their role models were sports figures. Additionally, as age increased, so did the proportion of sports figure role models. Interestingly, 95% of those boys had no interest in pursuing careers in sports. Also occurring during adolescence is a general trend for African Americans to disidentify with education and demonstrate increased drop-out rates (Griffin, 2002). Thus, these males are not interested in following the path of their role models, nor are they furthering their education. Perhaps role models play another role for them. For instance, perhaps the Black students are basking in the reflected glory (BIRG) of the sports figures (Kemmelmeier & Oyserman, 2001). Future research should address the characteristics of Black role models, as well as their purpose among those who hold role models. Of particular importance is determining what type of role models, if any, are helpful in alleviating stereotype threat affects for Black students, and at what age are they most useful.

Education. Anecdotal evidence suggests that many college educators hear from their students that they are auditory, visual, or even tactile learners; and because the professor's teaching style does not match their learning style, they simply cannot excel in the class. A student who has been informed that he is a tactile learner may experience frustration on a difficult test and attribute his frustration to a teaching method that does not match his learning style. It may be that some students are unintentionally falling victim to stereotype threat effects by ruminating about learning styles in the same way that a person ruminates about negative stereotypes about a social group. A second possibility is that informing students about their learning styles might actually be helpful to the student, such that frustration on a test can be attributed to poor teaching rather than to the self. Future research will examine the advantages and disadvantages of informing students of their personal learning styles and a possible link to stereotype threat effects.

Similarly, college educators may have also come across a number of students with diagnoses of learning disabilities. Through stereotype threat, learning disabilities may negatively impact the performance of a student beyond its boundaries. Unlike students aware of their learning styles, when students with learning disabilities experience frustration with a test, they may be less likely to attribute the frustration to the teacher and more likely to take responsibility. This may lead the student to perform more poorly on an important test. Future research should examine whether performance decrements associated with learning disabilities might be alleviated through techniques similar to those used to alleviate stereotype threat.

Footnote

[1] Impossible answers were inappropriate responses to math items that should not have occurred as the program was designed, but were recorded due to programming error. For instance, the response 'six' would be considered an impossible answer on a scale that ranged from one to five.

Appendix A

Packet A – Unrated Items

$\sqrt{x\sqrt{x\sqrt{x}}} =$

(A) $x^{7/8}$

(B) $x^{7/4}$

(C) $x^{15/16}$

(D) $x^{3/4}$

(E) $x^{15/8}$

1.

If $b \neq 0$ and $ab = \frac{b}{4}$, then $a =$

(A) $\frac{1}{8}$

(B) $\frac{1}{4}$

(C) $\frac{1}{2}$

(D) 1

(E) 4

2.

If $\sqrt{ab} = 3$, $b = c^3$, and $c = 3$, what is the value of $\frac{1}{a}$?

(A) $\frac{1}{3}$

(B) 1

(C) 3

(D) 9

(E) 27

3.

If $4p^2 = 36$ and $36 > 5q$, which of the following must be true?

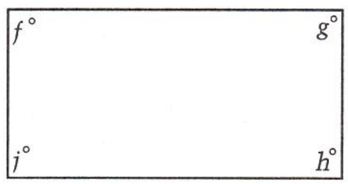
(A) $p^2 > 5q$
(B) $p^2 = 5q$
(C) $4p^2 > 5q$
(D) $4p^2 = 5q$
(E) $4p^2 < 5q$

4.

It takes a bus anywhere from 7 minutes to 10 minutes to travel from Town A to Town B. It takes the bus anywhere from 16 minutes to 24 minutes to travel from Town B to Town C. What are the least and greatest total travel times for a bus that travels from Town A to Town B and then from Town B to Town C? (Disregard the time the bus could be standing still in Town B.)

(A) 7 minutes and 24 minutes
(B) 10 minutes and 24 minutes
(C) 23 minutes and 26 minutes
(D) 23 minutes and 34 minutes
(E) 26 minutes and 40 minutes

5.



Note: Figure not drawn to scale

In the figure above, if $h = 20$, $g = 4h$, and $f = 2g$, what is the value of j?

(A) 60
(B) 80
(C) 90
(D) 100
(E) 260

6.

The sum of 7 numbers is greater than 140 and less than 210. Which of the following could be the average (arithmetic mean) of the numbers?

(A) 5
(B) 12
(C) 17
(D) 20
(E) 28

7.

By what number must the number 3.475817 be multiplied in order to obtain the number 34,758.17?

(A) 100
(B) 1,000
(C) 10,000
(D) 100,000
(E) 1,000,000

8.

If $x < 0 < y < 1$, which of the following CANNOT be true?

I. $xy = -\frac{1}{4}$

II. $\frac{x}{y} = -1$

III. $x + y > 1$

(A) I only
(B) II only
(C) III only
(D) I and II only
(E) I, II, and III

9.

If $7d < 4r$ and $4r < 8p$, which of the following must be true?

(A) $7d < 8p$
(B) $8p < 7d$
(C) $p < d$
(D) $r = 2p$
(E) $2r = p$

10.

Number of students	Grade
18	97
13	78
18	67
7	54
5	46

For a university class of 61 students, the table above shows the number of students receiving each grade on the mid-term exam. What is the median of those 61 scores?

(A) 97
(B) 78
(C) 67
(D) 54
(E) 46

11.

If a and b are positive integers and the ratio of $a + 1$ to $a + 2$ is the same as the ratio of $b + 3$ to $b + 4$, which of the following must be true?

I. $a = 4$
II. $b = 2$
III. $a - b = 2$

(A) I only
(B) II only
(C) III only
(D) I and II only
(E) I, II, and III

12.

Location	Number of Boxes	Number of Toys in Each Box
Basement	4	10
Garage	2	6
Attic	5	7

The chart above shows the location of all the toys stored at Joe's house. According to the chart, what is the total number of toys at Joe's house?

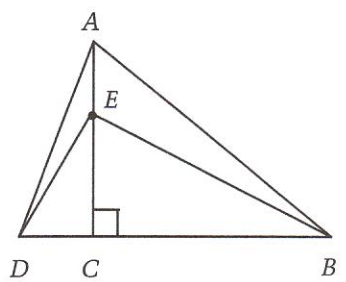
(A) 23
(B) 40
(C) 57
(D) 87
(E) 123

13.

If the manager of a store adds 50 lamps to its current inventory, the resulting total number of lamps is the same as three-halves of the current inventory. If the manager wanted to increase the current inventory by 40%, what would his new inventory of lamps be?

(A) 150
(B) 140
(C) 100
(D) 75
(E) 40

14.



Note: Figure not drawn to scale.

In the figure above, the length of *CB* is *t* percent less than the length of *DB*, and the length of *EC* is *t* percent less than the length of *AC*. If the area of triangle *EBC* is 16 percent of the area of triangle *ABD*, then what is the value of *t*?

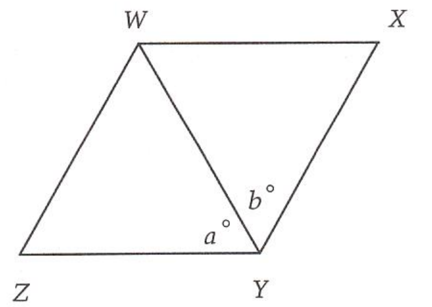
(A) 4
(B) 16
(C) 25
(D) 51
(E) 60

15.

If the average (arithmetic mean) of 8 numbers is greater than 10 and less than 12, which of the following could be the sum of the 8 numbers?

(A) 70
(B) 80
(C) 90
(D) 100
(E) 110

16.



In the figure above, $WX = XY = YZ = ZW = WY$. What is the value of $a + b$?

(A) 60
(B) 100
(C) 105
(D) 120
(E) 135

17.

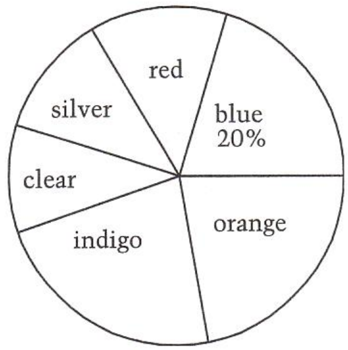
---

1. Select a number that is greater than 40 and less than 200.
2. Divide the number arrived at in the previous step by 20.
3. Find the smallest integer that is greater than or equal to the number arrived at in the previous step.
4. Subtract 5 from the number arrived at in the previous step.
5. Print the number that results.

Which of the following numbers could be printed in step 5?

(A) 17
(B) 12
(C) 1.7
(D) –1
(E) –7

18.



A certain necklace is made up of beads of the following colors: red, blue, orange, indigo, clear, and silver. The necklace contains 120 beads. According to the pie chart above, how many beads of the necklace are NOT blue?

(A) 20
(B) 24
(C) 92
(D) 96
(E) 100

19.

How much greater than the value of $3x - 7$ is the value of $3x + 5$?

(A) 12
(B) 10
(C) 7
(D) 5
(E) 2

20.

---

1. Select a number that is greater than 40 and less than 200.
2. Divide the number arrived at in the previous step by 20.
3. Find the smallest integer that is greater than or equal to the number arrived at in the previous step.
4. Subtract 5 from the number arrived at in the previous step.
5. Print the number that results.

If 150 is the number chosen in step 1, then what number will be printed in step 5?

(A) 2.5
(B) 3
(C) 7.5
(D) 8
(E) 70

21.

The hour hand of a watch rotates 30 degrees every hour. How many complete rotations does the hour hand make in 6 days?

(A) 18
(B) 12
(C) 10
(D) 8
(E) 6

22.

The Environment Club receives a certain amount of money from the school to host a teach-in. They budget 40% for a guest speaker, 25% for books, 20% for use of the auditorium, and the remainder for lunch. If the club plans to spend $90 on lunch for the participants, how much do they plan to spend on the guest speaker?

(A) $40
(B) $90
(C) $120
(D) $240
(E) $600

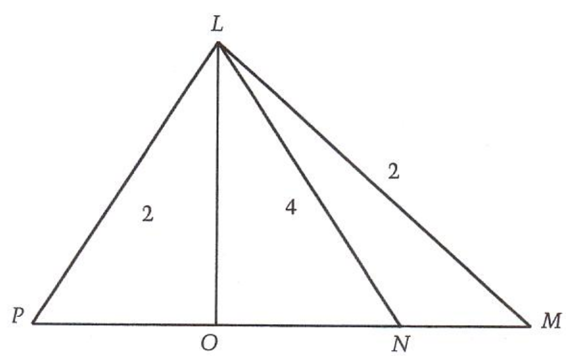
23.

1. Select a number that is greater than 40 and less than 200.
2. Divide the number arrived at in the previous step by 20.
3. Find the smallest integer that is greater than or equal to the number arrived at in the previous step.
4. Subtract 5 from the number arrived at in the previous step.
5. Print the number that results.

When the number 112 is selected in step 1, the number printed in step 5 is b. When the number a is selected in step 1, the number b is printed in step 5. What is the greatest possible value of a?

(A) 112
(B) 114
(C) 120
(D) 123
(E) 134

24.



Note: Figure not drawn to scale.

Which of the following is a triangle whose angles have degree measures of 30, 60, and 90?

(A) ΔLMN
(B) ΔLMP
(C) ΔLOP
(D) ΔLMO
(E) ΔLNO

25.

Which of the following statements expresses the statement "When z is decreased by 3, the result is twice the square of the sum of y and 4"?

(A) $z - 3 = 2(y + 4)^2$
(B) $z - 3 = 2(y^2 + 4^2)$
(C) $z = 2(y + 4)^2 - 3$
(D) $z - 3 = 2(y^2 + 4)$
(E) $z + 3 = 2(y + 4)^2$

26.

Which of the following expressions can be negative?

(A) $\frac{x^2}{2}$

(B) $\frac{3}{1 + x^2}$

(C) $4x^3$

(D) $(x^3)^2$

(E) $x(x^3 + x^5)$

27.

Line ℓ passes through the point $(-1, 2)$. Which of the following CANNOT be the equation of line ℓ?

(A) $y = 1 - x$
(B) $y = x + 1$
(C) $x = -1$
(D) $y = x + 3$
(E) $y = 2$

28.

The Earth makes one complete rotation about its axis every 24 hours. Assuming it rotates at a constant rate, through how many degrees would Goannaville, Australia rotate from 1:00 p.m. on January 2 to 4:00 p.m. on January 3?

(A) 202°
(B) 250°
(C) 350°
(D) 363°
(E) 405°

29.

For all numbers c and d, the symbol # is defined by $c \# d = (c + 1)(d - 1)$. What is the value of $(5 \# 2) \# (6 \# 4)$?

(A) 35
(B) 72
(C) 108
(D) 110
(E) 140

30.

Packet B – Unrated Math Items

If S is the set of all numbers between –3.5 and 3.5, inclusive, T is the set of all prime numbers, and U is the set of all positive integers, then the intersection of S, T, and U contains how many elements?

(A) 0
(B) 1
(C) 2
(D) 3
(E) More than 3

1.

If p is an integer greater than 1, such that p divided by 4 yields a remainder of 0, which of the following could be a prime number?

(A) $\frac{p}{4}$

(B) $2\sqrt{p}$

(C) $\frac{p}{3}$

(D) p

(E) $2p$

2.

Which of the following expressions must be positive for all values of a and b?

(A) $a + b$
(B) $a^2 - b^2 + 10$
(C) $a^2 + b^2 + 1$
(D) $a^3 + b^3 + 16$
(E) $a^4 + b^2 + a^2$

3.

If the length of one side of a triangle is 5, which of the following could be the perimeter of the triangle?

(A) 11
(B) 10
(C) 9
(D) 8
(E) 7

4.

Which of the following expressions is equal to 3^8 when $y = 3^5$?

(A) $\frac{y}{3}$

(B) $9y^2$

(C) $\frac{y^2}{3}$

(D) $\frac{y^2}{9}$

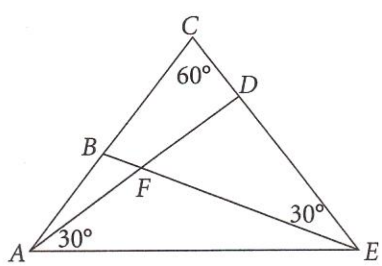
(E) $\frac{y^3}{27}$

5.

The ratio of the areas of three circles is 1 : 4 : 8, and the radius of the smallest circle is a positive integer. If the sum of the lengths of the diameters of the three circles is x, which of the following is a possible value of x?

(A) 12
(B) $8 + 4\sqrt{3}$
(C) 15
(D) $12 + 4\sqrt{2}$
(E) $18 + 12\sqrt{2}$

6.



Note: Figure not drawn to scale.

In triangle ACE above, AD and BE are line segments. Which of the following is NOT a right triangle?

(A) ABF
(B) ACD
(C) ADE
(D) AFE
(E) BCE

7.

For any odd integer x, where $x < 0$, how many negative, even integers are greater than x?

(A) $-x - 2$

(B) $\frac{-x}{2}$

(C) $\frac{-x-1}{2}$

(D) $x + 4$

(E) $\frac{-x-2}{2}$

8.

The initial number of elements in a certain set is p, where $p > 0$. If the number of elements in the set doubles every hour, which of the following represents the total number of elements in the set after exactly 24 hours?

(A) $24p$
(B) $48p$
(C) $2p^{24}$
(D) $(2p)^{24}$
(E) $(2^{24})p$

9.

A particular slot machine has three rotating wheels, called wheels A, B, and C. Each wheel displays the following pictures: a rose, a pen, a waterfall, an apple, a candle, a dollar sign, and an emerald. The machine awards a cash prize to a player whenever wheels A and B land on the same picture and wheel C lands on a candle. Assuming that for each wheel there is an equal probability of landing on each picture, what is the probability that the player will win a cash prize?

(A) $\frac{1}{343}$

(B) $\frac{1}{49}$

(C) $\frac{1}{21}$

(D) $\frac{1}{7}$

(E) $\frac{3}{7}$

10.

11. The absolute value of a certain integer is greater than 3 and less than 6. Which of the following could NOT be 2 less than the integer?

(A) −7
(B) −6
(C) 2
(D) 3
(E) 4

---

12. What part of three-fourths is one-tenth?

(A) $\frac{1}{8}$
(B) $\frac{15}{2}$
(C) $\frac{2}{15}$
(D) $\frac{3}{40}$
(E) None of these

---

13. A runner takes nine seconds to run a distance of 132 feet. What is the runner's speed in miles per hour?

(A) 9
(B) 10
(C) 11
(D) 12
(E) 13

---

14. $\frac{x-y}{x+y} - \frac{x+y}{x-y} =$

(A) $\frac{(4xy)}{\left(x^2-y^2\right)}$
(B) $\frac{(-4xy)}{\left(x^2-y^2\right)}$
(C) 0
(D) $\frac{(-2xy)}{\left(x^2-y^2\right)}$
(E) −1

$3(a-1) = 7(a+2)$. Find a.

(A) $\frac{17}{4}$

(B) $-\frac{17}{4}$

(C) $-\frac{14}{4}$

(D) $-\frac{4}{3}$

(E) $-\frac{3}{4}$

15.

What is the value of the following expression: $\dfrac{1}{1+\dfrac{1}{1+\frac{1}{4}}}$?

(A) $\frac{9}{5}$

(B) $\frac{5}{9}$

(C) $\frac{1}{2}$

(D) 2

(E) 4

16.

. Tickets for a particular concert cost $5 each if purchased in advance and $7 each if bought at the box office on the day of the concert. For this particular concert, 1,200 tickets were sold and the receipts were $6,700. How many tickets were bought at the box office on the day of the concert?

(A) 500

(B) 700

(C) 600

(D) 350

(E) 200

17.

. A counting number with exactly 2 different factors is called a prime number. Which of the following pairs of numbers are consecutive prime numbers?

(A) 27 and 29

(B) 31 and 33

(C) 35 and 37

(D) 37 and 29

(E) 41 and 43

18.

If $a \times b = 6a - 2bx$ and $9 \times 6 = 6$, then $x =$

(A) 2.

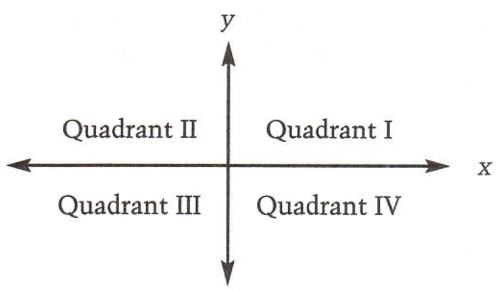
(B) 0.

(C) 1.

(D) 4.

(E) 3.

19.



If the product of the x-coordinate and y-coordinate of a point is 20, in which quadrant must that point lie?

(A) I
(B) II
(C) III
(D) IV
(E) It cannot be determined from the information given.

20.

Find the value of x in $2x + 12 = 3x + 9$.

(A) 1

(B) 2

(C) 3

(D) 4

(E) 5

21.

---

35 is 7% of what quantity?

(A) 2.45

(B) 5

(C) 245

(D) 50

(E) 500

22.

Find the median for the following set of numbers: 16, 22, 18, 21, 17, 21, 19, and 21.

(A) 21.0
(B) 20.0
(C) 22.0
(D) 19.0
(E) 19.4

23.

Tom received 89, 94, 86, and 96 on the first four algebra tests. What grade must he receive on his last test to have an average of 92?

(A) 92
(B) 94
(C) 91
(D) 95
(E) 96

24.

What is the median of the following group of scores?

27, 27, 26, 26, 26, 26, 18, 13, 36, 36, 30, 30, 30, 27, 29

(A) 30
(B) 26
(C) 25.4
(D) 27
(E) 36

25.

If $2a + 2b = 1$, and $6a - 2b = 5$, which of the following statements is true?

(A) $3a - b = 5$
(B) $a + b > 3a - b$
(C) $a + b = -2$
(D) $a + b < 3a - b$
(E) $a + b = -1$

26.

If n and k are even integers, which of the following is an even integer?

(A) $n + k + 1$
(B) $(n - 1)(k + 1)$
(C) $2(n + k + 1)$
(D) $(n - 3)(k + 1)$
(E) $2(n + k) + 1$

27.

In a class of 40 students, 30 speak French and 20 speak German. What is the lowest possible number of students who speak both languages?

(A) 5
(B) 20
(C) 15
(D) 10
(E) 30

28.

If $z > 0$, $x = z^2 + 3y$, and $y - 1 = z^2$, what is z in terms of x?

(A) $\sqrt{\frac{x}{2}}$

(B) $\sqrt{\frac{3+x}{2}}$

(C) $\sqrt{\frac{3-x}{2}}$

(D) $\sqrt{\frac{x+3}{4}}$

(E) $\sqrt{\frac{x-3}{4}}$

29.

Peter has five rulers of 30 cm each and three of 20 cm each. What is the average length of Peter's rulers?

(A) 25
(B) 27
(C) 23
(D) 26.25
(E) 27.25

30.

1. After taking four tests, Joan has an average grade of 79 points. What grade must she get on her fifth test to achieve an 83 point average?

(A) 83
(B) 86
(C) 87
(D) 95
(E) 99

---

2. What is t equal to if $A = P(1 + rt)$?

(A) $A - P - Pr$
(B) $\frac{(A+P)}{Pr}$
(C) $\frac{A}{P-r}$
(D) $\frac{(A-P)}{Pr}$
(E) None of these

---

3. $(5x - 3)(4x - 6) =$

(A) $20x^2 - 42x + 18$
(B) $20x^2 - 18$
(C) $20x^2 - 12x - 18$
(D) $30x^2 - 18$
(E) $18x^2 - 30x - 24$

---

4. The solution of the equation $4 - 5(2y + 4) = 4$ is

(A) $-\frac{2}{5}$.
(B) 8.
(C) 4.
(D) –2.
(E) none of these.

---

5. If $6x + 12 = 5$, then the value of $(x + 2)$ is

(A) $-{}^{19}/_{6}$.
(B) $-1{}^{1}/_{6}$.
(C) ${}^{5}/_{6}$.
(D) $3{}^{1}/_{6}$.
(E) $1{}^{1}/_{6}$.

If $V = \pi b^2\left(r - \frac{b}{3}\right)$, then r is equal to

(A) $\frac{V}{\pi b^2} + \frac{b}{3}$.

(B) $\frac{V}{\pi b^2} + \frac{b}{3\pi}$.

(C) $\frac{V}{\pi b^2} + 3b$.

(D) $V + \frac{b}{3}$.

(E) $V + \frac{\pi b}{3}$.

6.

. The value of B in the equation $a = \left(\frac{h}{2}\right)(B + b)$ is

(A) $\frac{(2a - b)}{h}$.

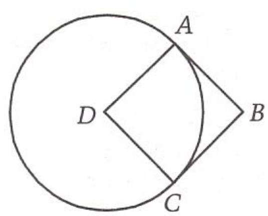
(B) $\frac{2h}{a - b}$.

(C) $2a - b$.

(D) $\frac{2a}{h} - b$.

(E) none of these.

7.



In the figure above, D is the center of the circle, and the perimeter of square $ABCD$ is 24. What is the area of the entire figure?

(A) $36 + 18\pi$
(B) $18 + 27\pi$
(C) $36 + 27\pi$
(D) $27 + 36\pi$
(E) $36 + 54\pi$

8.

Two pounds of pears and one pound of peaches cost $1.40. Three pounds of pears and two pounds of peaches cost $2.40. How much is the combined cost of one pound of pears and one pound of peaches?

(A) $2.00
(B) $1.50
(C) $1.60
(D) $.80
(E) $1.00

9.

If a triangle of base 6 units has the same area as a circle of radius 6 units, what is the altitude of the triangle?

(A) π
(B) 3π
(C) 6π
(D) 12π
(E) 36π

10.

$4\frac{1}{3} - 1\frac{5}{6} =$

(A) $3\frac{2}{3}$
(B) $2\frac{1}{2}$
(C) $3\frac{1}{2}$
(D) $2\frac{1}{6}$
(E) None of these

11.

If the measures of the three angles of a triangle are $(3x + 15)°$, $(5x - 15)°$, and $(2x + 30)°$, what is the measure of each angle?

(A) 75°
(B) 60°
(C) 45°
(D) 25°
(E) 15°

12.

13. If $0 < a < 1$ and $b > 1$, which is the largest value?

(A) $\frac{a}{b}$

(B) $\frac{b}{a}$

(C) $\left(\frac{a}{b}\right)^2$

(D) $\left(\frac{b}{a}\right)^2$

(E) Cannot be determined

---

14. The most economical price among the following prices is

(A) 10 oz. for 16¢.

(B) 2 oz. for 3¢.

(C) 4 oz. for 7¢.

(D) 20 oz. for 34¢.

(E) 8 oz. for 13¢.

---

15. A used car dealer reduced the price of all the cars on his lot by \$300. If a car was originally priced at \$1,195, what percent (to the nearest tenth) is the markdown of the sale price?

(A) 29.3%

(B) 8.4%

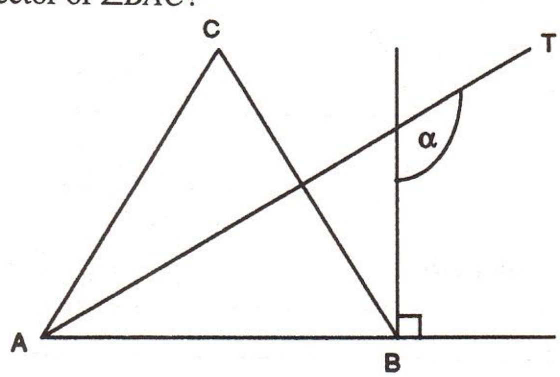
(C) 37.7%

(D) 25.1%

(E) 33.5%

---

16. In the figure, *ABC* is an equilateral triangle. What is the value of α if *AT* is the bisector of $\angle BAC$?



(A) 60°

(B) 90°

(C) 120°

(D) 45°

(E) 135°

17. Brand A drink contains 30 percent orange juice by volume. Brand B drink contains 40 percent orange juice by volume. Which of the following expressions gives the percent of orange juice in a mixture of x gallons of brand A drink, y gallons of brand B drink, and z gallons of water?

(A) $\frac{x+y}{x+y+z}\%$

(B) $\frac{40y}{x+y+z}\%$

(C) $\frac{30x+40y}{x+y}\%$

(D) $\frac{30x+40y}{x+y+z}\%$

(E) $\frac{30x+40y+z}{x+y+z}\%$

---

18. The sum of three consecutive odd integers is always divisible by (I) 2, (II) 3, (III) 5, or (IV) 6.

(A) I only

(B) II only

(C) III only

(D) I and III only

(E) IV only

---

19. A given cube has a surface area of 96 square feet. What is the volume of the cube in cubic feet?

(A) 16

(B) 36

(C) 64

(D) 96

(E) 216

---

20. The number missing in the series, 2, 6, 12, 20, x, 42, 56,

(A) 36.

(B) 24.

(C) 30.

(D) 38.

(E) 40.

The side of a square increases 10% and the area increases $5.25(\text{ft})^2$. What was the original value of the side of the square?

(A) 3 ft

(B) 2 ft

(C) 1 ft

(D) 4 ft

(E) 5 ft

21.

For non-zero numbers p, q, r, and s, $p/q = r/s$. Which of the following statements is true?

(A) $\frac{p+q}{p} = \frac{r+s}{r}$

(B) $\frac{p}{r} = \frac{q}{r}$

(C) $\frac{r}{s} = \frac{q}{p}$

(D) $\frac{p-q}{q} = \frac{s-r}{r}$

(E) $\frac{q}{p-q} = \frac{r}{s-r}$

22.

What is the reciprocal of $1 : {}^{8}/_{3}$?

(A) $\frac{8}{3}$

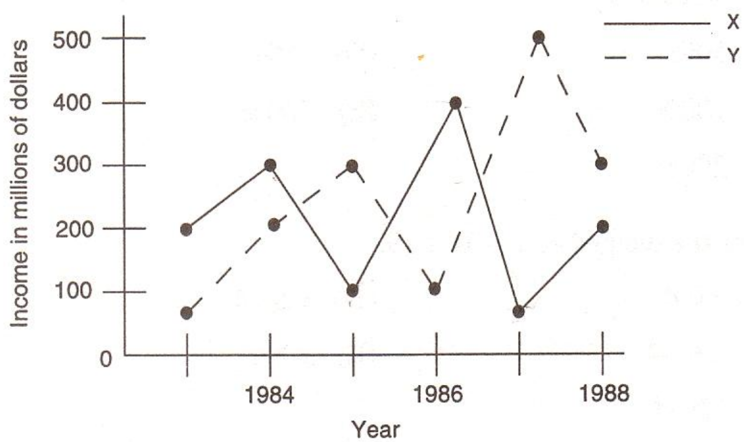
(B) $\frac{3}{8}$

(C) $1\frac{8}{3}$

(D) $1\frac{3}{8}$

(E) $1+\frac{8}{3}$

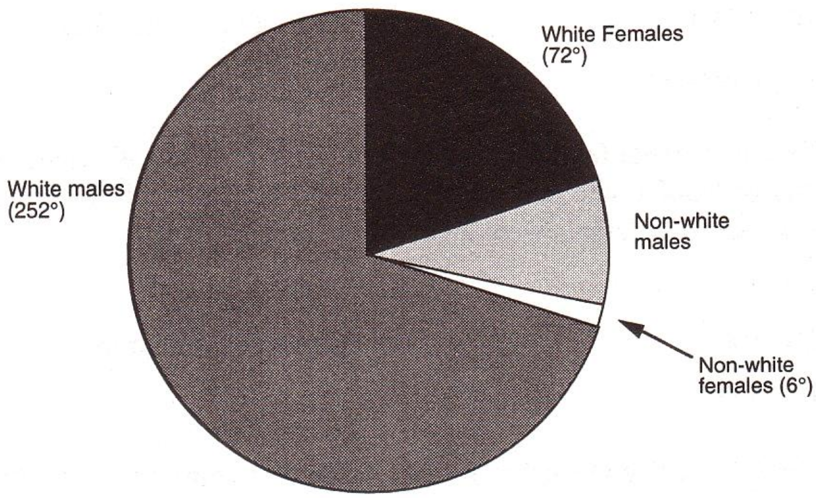
23.



24. What was the average income (in millions) of Company X during the years 1983 to 1986?

(A) 150
(B) 200
(C) 250
(D) 300
(E) 400

Portion of Doctoral Degrees in the Mathematical Sciences
Awarded to U.S. Citizens in 1986



What percent of the Ph.D. degrees were awarded in 1986 to non-white males?

(A) 30
(B) $8\frac{1}{3}$
(C) $4\frac{1}{6}$
(D) 20
(E) None of these

25.

One number is 2 more than 3 times another. Their sum is 22. Find the numbers.

(A) 8, 14
(B) 2, 20
(C) 5, 17
(D) 4, 18
(E) 10, 12

26.

Solve the inequality $7 - 3x \leq 19$.

(A) $x = 4$
(B) $x = -4$
(C) $x \geq -4$
(D) $x \leq -4$
(E) $x \geq 4$

27.

Which of the following equations can be used to find a number x, if the difference between the square of this number and 21 is the same as the product of 4 times the number?

(A) $x - 21 = 4x$

(B) $x^2 - 21 = 4x$

(C) $x^2 = 21 - 4x$

(D) $x + 4x^2 = 21$

(E) $x^2 + 21 = 4x$

28.

I went to Lucky Duck Casino and in the first game I lost one-third of my money; in the second game I lost half of the rest. If I still have $1,000, how much money did I have when I arrived at the casino?

(A) $1,000

(B) $2,000

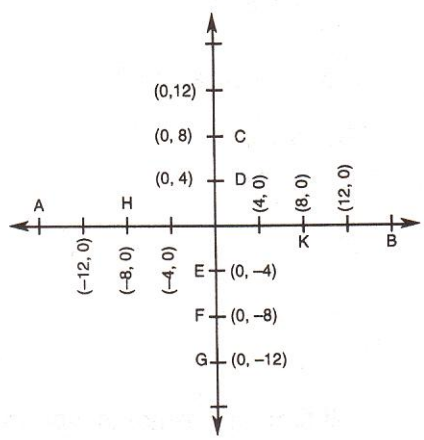
(C) $3,000

(D) $6,000

(E) $12,000

29.

Line $\overleftrightarrow{AB}$ is the perpendicular bisector of segment $\overline{CP}$ (P is not shown). Then P is the same as which of the following points?



(A) G

(B) F

(C) H

(D) K

(E) D

30.

The length of a rectangle is $6L$ and the width is $4W$. What is the perimeter?

(A) $12L + 8W$

(B) $12L^2 + 8W^2$

(C) $6L + 4W$

(D) $20LW$

(E) $24LW$

1.

A truck contains 150 small packages, some weighing 1 kg each and some weighing 2 kg each. How many packages weighing 2 kg each are in the truck if the total weight of all the packages is 264 kg?

(A) 36

(B) 52

(C) 88

(D) 124

(E) 114

2.

What is the factorization of $x^2 + ax - 2x - 2a$?

(A) $(x + 2)(x - a)$

(B) $(x - 2)(x + a)$

(C) $(x + 2)(x + a)$

(D) $(x - 2)(x - a)$

(E) None of these

3.

What is the length of side BC?

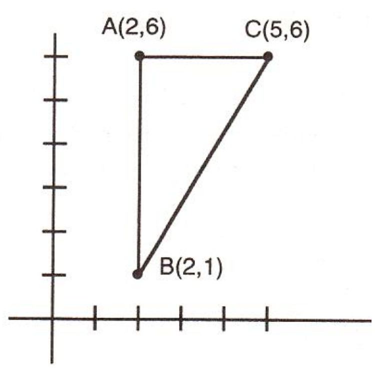
(A) 3

(B) 5

(C) $\sqrt{34}$

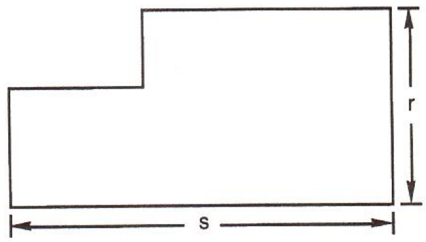
(D) 7

(E) None of these



4.

In the figure shown, all segments meet at right angles. Find the figure's perimeter in terms of r and s.



(A) $r + s$

(B) $2r + s$

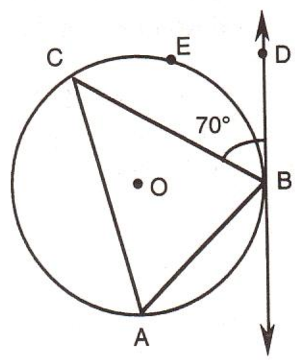
(C) $2s + r$

(D) $r^2 + s^2$

(E) $2r + 2s$

5.

The measure of an inscribed angle is equal to one-half the measure of its inscribed arc. In the figure shown, ΔABC is inscribed in circle O, and $\overleftrightarrow{BD}$ is tangent to the circle at point B. If the measure of $\angle CBD$ is 70°, what is the measure of $\angle BAC$?



(A) 110°

(B) 70°

(C) 140°

(D) 35°

(E) 40°

6.

Which of the following statements are true, if

$$x + y + z = 10$$

$$y \geq 5$$

$$4 \geq z \geq 3$$

I. $x < z$

II. $x > y$

III. $x + z \leq y$

(A) I only

(B) II only

(C) III only

(D) I and III only

(E) I, II, and III

7.

If the length of a rectangle is increased by 30% and the width is decreased by 20%, then the area is increased by

(A) 10%.

(B) 5%.

(C) 4%.

(D) 20%.

(E) 25%.

8.

A wheel with a diameter of 3 feet makes a revolution every 2 minutes. How many feet will the wheel travel in 30 minutes?

(A) 3π

(B) 6π

(C) 45π

(D) 30π

(E) 15π

9.

Jim is twice as old as Susan. If Jim were 4 years younger and Susan were 3 years older, their ages would differ by 12 years. What is the sum of their ages?

(A) 19

(B) 42

(C) 56

(D) 57

(E) None of these

10.

In the Carco Auto Factory, robots assemble cars. If 3 robots assemble 17 cars in 10 minutes, how many cars can 14 robots assemble in 45 minutes if all robots work at the same rate all the time?

(A) 357

(B) 340

(C) 705

(D) 150

(E) 272

11.

Given $\frac{(\alpha + x) + y}{x + y} = \frac{\beta + y}{y}$, $\frac{x}{y} =$

(A) $\frac{\alpha}{\beta}$

(B) $\frac{\beta}{\alpha}$

(C) $\frac{\beta}{\alpha - 1}$

(D) $\frac{\alpha}{\beta} - 1$

(E) 1

12.

If n is an integer, which of the following represents an odd number?

(A) $2n + 3$

(B) $2n$

(C) $2n + 2$

(D) $3n$

(E) $n + 1$

13.

If R, S, and Q can wallpaper a house in 8 hours and R and S can do it in 12 hours, how long will it take Q alone to wallpaper the house?

(A) 12 hours

(B) 24 hours

(C) 8 hours

(D) 20 hours

(E) 28 hours

14.

15. $\left[\dfrac{.0003 \times 9 \times 10^{-1}}{18 \times 10^{-4}}\right]^{-1} =$

(A) $\dfrac{20}{3}$

(B) $\dfrac{3}{20}$

(C) $\dfrac{3}{2}$

(D) $\dfrac{2}{3}$

(E) $\dfrac{19}{3}$

---

16. If n is the first of three consecutive odd numbers, which of the following represents the sum of the three numbers?

(A) $n + 2$

(B) $n + 4$

(C) $n + 6$

(D) $3n + 6$

(E) $6(3n)$

---

17. A waitress's income consists of her salary and tips. Her salary is $150 a week. During one week that included a holiday, her tips were $5/4$ of her salary. What fraction of her income for the week came from tips?

(A) $\dfrac{5}{8}$

(B) $\dfrac{5}{4}$

(C) $\dfrac{4}{9}$

(D) $\dfrac{1}{2}$

(E) $\dfrac{5}{9}$

---

18. Joe and Jim together have 14 marbles. Jim and Tim together have 10 marbles. Joe and Tim together have 12 marbles. What is the maximum number of marbles that any one of these may have?

(A) 7

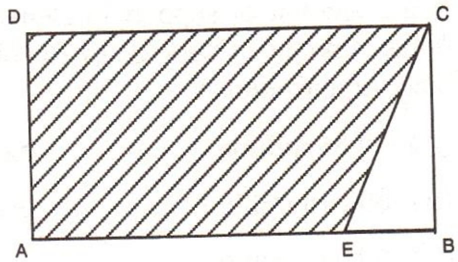
(B) 8

(C) 9

(D) 10

(E) 11

If the length of segment $\overline{EB}$; base of triangle EBC, is equal to $^1/_4$ the length of segment $\overline{AB}$ ($\overline{AB}$ is the length of rectangle $ABCD$), and the area of triangle EBC is 12 square units, find the area of the shaded region.



(A) 24 square units

(B) 96 square units

(C) 84 square units

(D) 72 square units

(E) 120 square units

19.

Which of the following options is correct?

(A) $\alpha + \beta + \gamma = 180°$

(B) $\gamma - \alpha + 180° = \beta$

(C) $\alpha = \beta + \gamma$

(D) $\gamma = \alpha + \beta$

(E) $\alpha = 180° - \beta - \gamma$

20.

A postal truck leaves its station and heads for Chicago, averaging 40 mph. An error in the mailing schedule is spotted and 24 minutes after the truck leaves, a car is sent to overtake the truck. If the car averages 50 mph, how long will it take to catch the postal truck?

(A) 2.6 hours

(B) 3 hours

(C) 2 hours

(D) 1.5 hours

(E) 1.6 hours

21.

The quotient of $\frac{\left(x^2-5x+3\right)}{(x+2)}$ is

(A) $x-7+\frac{17}{(x+2)}$.

(B) $\frac{x-3+9}{(x+2)}$.

(C) $\frac{x-7-11}{(x+2)}$.

(D) $\frac{x-3-3}{(x+2)}$.

(E) $\frac{x+3-3}{(x+2)}$.

22.

The perimeter of a square inscribed in the circumference of radius R is

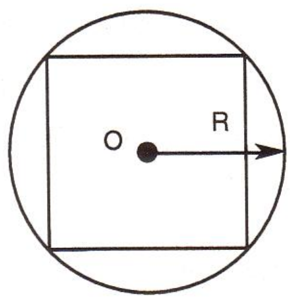
(A) $4R$.

(B) $8R$.

(C) $2R\sqrt{2}$.

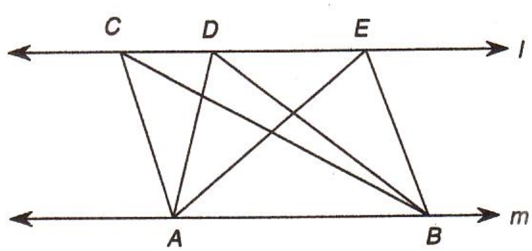
(D) $4R\sqrt{2}$.

(E) $8R\sqrt{2}$.



23.

In the following figure, line *l* is parallel to line *m*. If the area of ΔABC is 40 cm^2, what is the area of triangle ΔABD?



24.

Each of the integers h, m, and n is divisible by 3. Which of the following integers is always divisible by 9?

I. hm

II. $h + m$

III. $h + m + n$

(A) I only

(B) II only

(C) III only

(D) II and III only

(E) I, II, and III

25.

The range for $|5x - 1| \le 9$ is

(A) $-\frac{8}{5} \le x \le 2.$

(B) $-\frac{9}{5} \le x \le \frac{9}{5}.$

(C) $0 \le x \le \frac{1}{5}.$

(D) $-\frac{8}{5} \le x \le \frac{9}{5}.$

(E) $-\frac{8}{5} \ge x$ and $x \ge 2.$

26.

A doctor has 40 cc of 2% tincture of iodine. If the iodine is boiled, alcohol is evaporated away and the strength of tincture is raised. How much alcohol must be boiled away in order to raise the strength of the tincture to 8%?

(A) 34 cc

(B) 32 cc

(C) 30 cc

(D) 24 cc

(E) 10 cc

27.

If $3^x > 1$, then x

(A) $0 < x < 1$

(B) $x \ge 0$

(C) $x \ge 1$

(D) $x > 0$

(E) $x > 1$

28.

$\sqrt{75} - 3\sqrt{48} + \sqrt{147} =$

(A) $3\sqrt{3}$

(B) $7\sqrt{3}$

(C) 0

(D) 3

(E) $\sqrt{3}$

29.

An old picture has dimensions 33 inches by 24 inches. What one length must be cut from each dimension so that the ratio of the shorter side to the longer side is 2 : 3?

(A) 2 inches

(B) 6 inches

(C) 9 inches

(D) $10^1/_2$ inches

(E) 3 inches

30.

Appendix B
STATEMENT OF CONSENT – Fall 2005

I, the undersigned, do hereby give my informed consent to my participation in the **Key Lime Pie** Study. I have been informed about each of the following:

- The purpose of the study is to test new items for the quantitative portion of the Graduate Record Examination.
- The procedures of the study include answering a few questions about math items.
- The benefits of the study include the opportunity to be involved in psychological experiments like the ones I've learned about in class.
- The risks of the study are negligible. After the completion of the study, the experimenter will answer any questions that I may have about the procedures.
- I understand that I will receive credit for this experiment at its completion and I cannot receive credit for participation in the current experiment more than once.

I understand that I may withdraw at any time before or during the experiment at my option.

Recognizing the importance of avoiding bias in the results of this experiment, I agree not to discuss any of the details of the procedure with other participants. I understand that all of the research and evaluation materials will be confidentially maintained. The means used to maintain confidentiality are:

1. My data will be given a code number for research identification, and my name will be kept anonymous.
2. Data, along with consent forms, will be kept in a locked file cabinet.
3. Only the investigators will have access to my identification data.

I understand that if I have questions concerning the research, I can call the following persons:

Dana Gresky, Principal Investigator
Department of Psychology
257-7414

Laura Ten Eyck, Co-principal Investigator
Department of Psychology
257-7414

Dr. Don Dansereau
Chair, Department of Psychology
Human Subjects Committee
257-7410

Jan Fox, TCU Coordinator
Research and Sponsored Projects
257-7515

Dr. Charles Lord
Faculty Sponsor
257-7410

______________________________ ______________
Participant's Name (PLEASE PRINT) Date

______________________________ ______________
Participant's Signature Phone Number

______________________________ ______________
Participant's TCU Student ID# Professor

Appendix C

Domain Identification Measure (Smith & White, 2001)

Using the following scale, please indicate the number that best describes how much you agree with each of the statements below.

1	2	3	4	5
Strongly Disagree	**Moderately Disagree**	**Neither Disagree or Agree**	**Moderately Agree**	**Strongly Agree**

1. I learn things quickly in English classes. _______
2. Mathematics is one of my best subjects. _______
3. English is one of my best subjects. _______
4. I get good grades in English. _______
5. I have always done well in Math. _______
6. I'm hopeless in English classes. _______
7. I get good grades in Math. _______
8. I do badly in tests of Mathematics. _______

Please indicate the number that best describes you for each of the statements below using the following scale:

1	2	3	4	5
Not at all		**Somewhat**		**Very Much**

9. How much do you enjoy math-related subjects? _______
10. How much do you enjoy English-related subjects? _______
11. How likely would you be to take a job in a math related field? _______
12. How much is math to the sense of who you are? _______

Please indicate the number that best describes you for each of the statements below using the following scale:

1	2	3	4	5
Not at all		**Somewhat**		**Very Much**

13. How important is it to you to be good at math? _______
14. How important is it to you to be good at English? _______
15. Compared to other students, how good are you at math?
    a. Very Poor
    b. Poor
    c. About the same
    d. Better than average
    e. Excellent
16. Compared to other students, how good are you at English?
    a. Very Poor
    b. Poor
    c. About the same
    d. Better than average
    e. Excellent

Appendix D

STATEMENT OF CONSENT – Fall 2005

I, the undersigned, do hereby give my informed consent to my participation in the **Jelly Donut** Study. I have been informed about each of the following:

- The purpose of the study is to test new items for the quantitative portion of the Graduate Record Examination.
- The procedures of the study include answering a series of math questions.
- The benefits of the study include the opportunity to be involved in psychological experiments like the ones I've learned about in class.
- The risks of the study are negligible. After the completion of the study, the experimenter will answer any questions that I may have about the procedures.
- I understand that I will receive credit for this experiment at its completion and I cannot receive credit for participation in the current experiment more than once.

I understand that I may withdraw at any time before or during the experiment at my option.

Recognizing the importance of avoiding bias in the results of this experiment, I agree not to discuss any of the details of the procedure with other participants. I understand that all of the research and evaluation materials will be confidentially maintained. The means used to maintain confidentiality are:

4. My data will be given a code number for research identification, and my name will be kept anonymous.
5. Data, along with consent forms, will be kept in a locked file cabinet.
6. Only the investigators will have access to my identification data.

I understand that if I have questions concerning the research, I can call the following persons:

Dana Gresky, Principal Investigator
Department of Psychology
257-7414

Laura Ten Eyck, Co-principal Investigator
Department of Psychology
257-7414

Dr. Don Dansereau
Chair, Department of Psychology
Human Subjects Committee
257-7410

Jan Fox, TCU Coordinator
Research and Sponsored Projects
257-7515

Dr. Charles Lord
Faculty Sponsor
257-7410

________________________________ ________________

Participant's Name (PLEASE PRINT) Date

________________________________ ________________

Participant's Signature Phone Number

____________________ ____________________

Participant's TCU Student ID# Professor

Appendix E
PowerPoint Welcome Message

First screen:
"Welcome!"

Second screen:
"A little information before you begin… computer based-programs are new to the department."

Third screen:
"Some things may be confusing… or you may be unsure how to respond."

Fourth screen:
"If you have any questions…raise your hand and wait for the experimenter."

Fifth screen:
"We will be happy to assist you!"

Appendix F

PowerPoint Practice Math Test Instructions:

First screen:
"Your first task today will be to work on some practice math questions. We would like you to get acquainted with completing math questions presented in a computer format."

Second screen:
"You will be presented with 6 practice questions that each have 5 answer choices. After determining the correct answer, simply click on the box that corresponds to that answer choice."

Third screen:
"Pencils and some scratch paper have been provided, feel free to use them to work out the problems. When you are ready to begin the practice test, press continue below."

Appendix G
Experiment 1 Practice Test

Question 1

If the manager of a store adds 50 lamps to its current inventory, the resulting total number of lamps is the same as three-halves of the current inventory. If the manager wanted to increase the current inventory by 40%, what would his new inventory of lamps be?

(A) 150
(B) 140
(C) 100
(D) 75
(E) 40

Question 2

A wheel with a diameter of 3 feet makes a revolution every 2 minutes. How many feet will the wheel travel in 30 minutes?

(A) 3π
(B) 6π
(C) 45π
(D) 30π
(E) 15π

Question 3

The quotient of $\frac{(x^2 - 5x + 3)}{(x+2)}$ is

(A) $x - 7 + \frac{17}{(x+2)}$.
(B) $\frac{x-3+9}{(x+2)}$.
(C) $\frac{x-7-11}{(x+2)}$.
(D) $\frac{x-3-3}{(x+2)}$.
(E) $\frac{x+3-3}{(x+2)}$.

Question 4

$\sqrt{75} - 3\sqrt{48} + \sqrt{147} =$

(A) $3\sqrt{3}$

(B) $7\sqrt{3}$

(C) 0

(D) 3

(E) $\sqrt{3}$

Question 5

1. Select a number that is greater than 40 and less than 200.
2. Divide the number arrived at in the previous step by 20.
3. Find the smallest integer that is greater than or equal to the number arrived at in the previous step.
4. Subtract 5 from the number arrived at in the previous step.
5. Print the number that results.

When the number 112 is selected in step 1, the number printed in step 5 is b. When the number a is selected in step 1, the number b is printed in step 5. What is the greatest possible value of a?

(A) 112
(B) 114
(C) 120
(D) 123
(E) 134

Question 6

A doctor has 40 cc of 2% tincture of iodine. If the iodine is boiled, alcohol is evaporated away and the strength of tincture is raised. How much alcohol must be boiled away in order to raise the strength of the tincture to 8%?

(A) 34 cc

(B) 32 cc

(C) 30 cc

(D) 24 cc

(E) 10 cc

Appendix H
PowerPoint Transition from Pretest to GRE

First screen:
"Great Practice!"

Second screen"
"Now that you've gotten acquainted with the computer format…"

Third screen:
"…you'll be able to take the real test."

Appendix I
GRE Logo

GRADUATE RECORD EXAMINATIONS®

GRE®

QUANTITATIVE TEST

Appendix J
Punishment/Reward Instructions

Punishment Instructions

The quantitative reasoning section of the General Test is an **adaptive** test. It is **tailored to your performance** level and provides precise information about your abilities using fewer test questions than traditional paper-based tests.

At the start of the test, you are presented with test questions of middle difficulty. As you answer each question, the computer scores it and uses that information, as well as your responses to preceding questions and information about the test design, to determine the next question. As long as you answer correctly, you will typically be given questions of slightly increased difficulty. When you respond **incorrectly to several questions in a row,** you will typically be given **a couple of less difficult questions** as a way to give you some test items appropriate to your ability level.

Because the computer scores each question before selecting the next one, you must answer each question when it is presented. For this reason, once you answer a question and move on to another, you cannot go back and change your answer. The computer has already incorporated both your answer and the requirements of the test design into its selection of your next question.

Each computer-based test meets established specifications, including the types of questions asked and the subject matter presented. The statistical characteristics of the questions answered correctly and incorrectly, including the difficulty levels, are taken into account in the calculation of scores. Therefore, it is appropriate to compare scores of different test takers even though they received different questions.

If it has not already done so, in a moment, "continue" will appear in the bottom right-hand corner of the screen. When you are sure you understand the above directions, click "continue."

Reward Instructions

The quantitative reasoning section of the General Test is an **adaptive** test. It is **tailored to your performance** level and provides precise information about your abilities using fewer test questions than traditional paper-based tests.

At the start of the test, you are presented with test questions of middle difficulty. As you answer each question, the computer scores it and uses that information, as well as your responses to preceding questions and information about the test design, to determine the next question. As long as you answer correctly, you will typically be given questions of slightly increased difficulty. When you respond **correctly to several questions in a row,** you will typically be given **a couple of less difficult questions** as a **reward,** in the form of a refreshing rest, and then get back to items appropriate to your ability.

Because the computer scores each question before selecting the next one, you must answer each question when it is presented. For this reason, once you answer a question and move on to another, you cannot go back and change your answer. The computer has already incorporated both your answer and the requirements of the test design into its selection of your next question.

Each computer-based test meets established specifications, including the types of questions asked and the subject matter presented. The statistical characteristics of the questions answered correctly and incorrectly, including the difficulty levels, are taken into account in the calculation of scores. Therefore, it is appropriate to compare scores of different test takers even though they received different questions.

If it has not already done so, in a moment, “continue” will appear in the bottom right-hand corner of the screen. When you are sure you understand the above directions, click “continue.”

Appendix K
Experimenter Script

Before participants begin:

- *Make sure that all electronic devices are turned off – not on vibrate.*
- *Reiterate that they should raise their hand if they have a question.*
- *When everyone has arrived, they may begin (wait as long a 5 minutes past start time). If someone arrives late, inform them that they may reschedule on ExperimenTrak for a different session.*
- *Pass out informed consent forms. Say something like:*
    - "This is the informed consent form, it says that you will be participating in the Jelly Donut Study. There are no risks involved in this study. Hopefully, by the end, you will learn something about psychology. All of the information that you provide for us today will remain confidential. Please make sure to read over the form before signing it. After you have completed the form, you may go ahead and click 'continue' on your computer."

During experiment:

- *Experimenter may use the computer or study (or stare into space! ☺).*
- PLEASE *walk around the room every 5 minutes or so and make sure that nothing weird is happening.*

Debriefing Script

- For Jelly Donut Experiment 1
    - "The experiment that you just completed was on stereotype threat. Stereotype threat is a performance debilitating phenomenon that affects members of certain groups when they are reminded about the relevant stereotype. For today's experiment, we used a subtle reminder, that this is a test that 'has shown gender differences in the past' to remind women of the stereotype that they are not typically perceived as being good at math."
    - "During the experiment, you were told that the test items you received were based upon your performance. In actuality, there was only one test tat was administered to everyone. But some of you were told that when you received easy questions, it was because you had been doing well, and for some, you were told that the easy questions were a result of performing poorly."
    - "We expected that for women, being notified that they were expected to do poorly, and then receiving confirmation of that – by receiving easy questions – would hurt math performance on questions later in the test."
    - "We hope that it works!"
    - "You've got just 2 or 3 more questions to respond to before we can let you go, and then, you're done! So, turn back to your computers, answer the last couple of questions, and you're free to leave."

- For Jelly Donut Experiment 2
    - "The experiment that you just completed was on stereotype threat. Stereotype threat is a performance debilitating phenomenon that affects members of

certain groups when they are reminded about the relevant stereotype. For today's experiment, we used a subtle reminder, that this is a test that 'has shown gender differences in the past' to remind women of the stereotype that they are not typically perceived as being good at math."

- "During the experiment, there were two basic test formats. On one test, the majority of the questions were fairly easy, while 6 were quite difficult. On the other test, the majority of the questions were quite difficult, and it included the exact same 6 questions included on the easy test."
- "We expected that for women, being notified that they were expected to do poorly and laboring through a difficult test, would disrupt their performance on those 6 items that are shared by each test. But, for the women with the mostly easy test, the six math problems should be perceived as easier based upon their context."
- "We hope that it works!"
- "You've got just 2 or 3 more questions to respond to before we can let you go, and then, you're done! So, turn back to your computers, answer the last couple of questions, and you're free to leave."

Appendix L

Experiment 2 Practice test

Question 1

If S is the set of all numbers between –3.5 and 3.5, inclusive, T is the set of all prime numbers, and U is the set of all positive integers, then the intersection of S, T, and U contains how many elements?

(A) 0
(B) 1
(C) 2
(D) 3
(E) More than 3

Question 2

Two pounds of pears and one pound of peaches cost $1.40. Three pounds of pears and two pounds of peaches cost $2.40. How much is the combined cost of one pound of pears and one pound of peaches?

(A) $2.00
(B) $1.50
(C) $1.60
(D) $.80
(E) $1.00

Question 3

Which of the following statements are true, if

$$x + y + z = 10$$

$$y \geq 5$$

$$4 \geq z \geq 3$$

I. $x < z$

II. $x > y$

III. $x + z \leq y$

(A) I only

(B) II only

(C) III only

(D) I and III only

(E) I, II, and III

Question 4

If the length of a rectangle is increased by 30% and the width is decreased by 20%, then the area is increased by

(A) 10%.

(B) 5%.

(C) 4%.

(D) 20%.

(E) 25%.

Question 5

If R, S, and Q can wallpaper a house in 8 hours and R and S can do it in 12 hours, how long will it take Q alone to wallpaper the house?

(A) 12 hours

(B) 24 hours

(C) 8 hours

(D) 20 hours

(E) 28 hours

Question 6

If n is the first of three consecutive odd numbers, which of the following represents the sum of the three numbers?

(A) $n + 2$

(B) $n + 4$

(C) $n + 6$

(D) $3n + 6$

(E) $6(3n)$

Appendix M
Math Item Layout (Experiment 2)

Difficult in Difficult				Difficult Easy		
Diff	SD	Number		Diff	SD	Number
6.26	1.75	A1	1	3.52	1.86	A2
5.73	1.99	A14	2	2.15	1.57	A5
6.84	1.80	A15	3	2.94	1.84	A6
6.21	1.93	A24	4	2.63	1.70	A11
5.89	1.62	B5	5	3.05	1.71	B15
4.63	**1.92**	**A9**	**6**	**4.63**	**1.92**	**A9**
6.47	1.77	B6	7	3.68	1.79	B18
6.47	1.47	B29	8	2.21	1.58	B21
7.05	1.85	C17	9	1.94	1.17	B22
6.71	1.72	C21	10	2.47	1.21	B24
4.76	**1.60**	**C13**	**11**	**4.76**	**1.60**	**C13**
5.39	1.82	D9	12	2.05	1.68	B25
5.33	1.74	D10	13	2.47	1.47	B30
5.61	1.38	D19	14	2.35	1.37	C1
4.63	**1.60**	**B26**	**15**	**4.63**	**1.60**	**B26**
5.61	1.75	D21	16	2.19	1.68	C4
6.00	1.90	D22	17	1.82	1.63	C20
6.56	1.75	D23	18	3.17	1.65	D3
4.94	**1.87**	**B3**	**19**	**4.94**	**1.87**	**B3**
5.17	1.42	D26	20	2.50	1.65	D4
5.83	2.06	D29	21	2.89	1.27	D5
4.94	**1.95**	**B19**	**22**	**4.94**	**1.95**	**B19**

Note. Diff **= Mean level of difficulty as rated by participants in pretesting. *SD* = Standard deviation. *Number* = Math item denoted by test packet and item number. Easy items range from: 1.90 – 3.68 (1.78 point range); difficult items range from: 5.33 – 7.05 (1.72 point range); shared items range from: 4.63 – 4.94 (.31 point range).**

References

Ambady, N., Paik, S. K., Steele, J., Owen-Smith, A., & Mitchell, J. P. (2004). Deflecting negative self-relevant stereotype activation: The effects of individuation. *Journal of Experimental Social Psychology, 40,* 401-408.

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

Arndt, J., Greenberg, J., Schimel, J., Pyszczynski, T., & Solomon, S. (2002). To belong or not to belong, that is the question: Terror management and identification with gender and ethnicity. *Journal of Personality and Social Psychology, 83,* 26-43.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (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., Fried, C. B., & Good, C. (2001). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental and Social Psychology, 38,* 113-125.

Assibey-Mensah, G. O. (1997). Role models and youth development: Evidence and lessons from the perceptions of African American youth. *The Western Journal of Black Studies, 21,* 242-252.

Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173-1182.

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

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

Bond, C. F. (1982). Social facilitation: A self-presentational view. *Journal of Personality and Social Psychology, 42,* 1042-1050.

Bond, C. F., Atoum, A. O., & VanLeeuwen, M. D. (1996). Social impairment of complex learning in the wake of public embarrassment. *Basic and Applied Social Psychology, 18,* 31-44.

Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology, 40,* 247-255.

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

Brown, R. P., & Pinel, E. C. (2003). Sigma on my mind: Individual differences in the experience of stereotype threat, *Journal of Experimental Social Psychology, 39,* 626-633.

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology, 33,* 267-285.

Crocker, J. & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review, 96,* 608-630.

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, 6,* 588-594.

Croizet, J-C., Dutrévis, M., & Désert, M. (2002). Why do students holding non-prestigious high school degrees underachieve at the university? *Swiss Journal of Psychology, 61,* 167-175.

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28,* 1615-1628.

Educational Testing Service, Graduate Record Examination. (2005) Retrieved January 1, 2005, from www.gre.org.

Feather, N. T., & Simon, J. G. (1971). Causal attributions for success and failure in relation to expectations of success based upon selective or manipulative control. *Journal of Personality, 39,* 527-541.

Franz, C. M., Cuddy, A. J. C., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The race implicit association test as a stereotype threat experience. *Personality and Social Psychology Bulletin*, *30,* 1611-1624.

Gilmor, T. M. & Minton, H. L. (1974). Internal versus external attribution of task performance as a function of locus of control, initial confidence and success-failure outcome. *Journal of Personality, 42,* 159-174.

Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of women. *Personality and Social Psychology Bulletin, 28,* 659-670.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24,* 625-662.

Gresky, D. M., Ten Eyck, L. L., Lord, C. L., & McIntyre, R. B. (2005). Effects of salient multiple identities on women's performance under mathematics stereotype threat. *Sex Roles, 53,* 703-716.

Griffin, B. W. (2002). Academic disidentification, race, and high school dropouts. *High School Journal, 85,* 71-81.

Hedges, L. V. & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high scoring individuals. *Science, 269,* 41-45.

Hess, T. M. & Hinson, J. T. (2006). Age-related variation in the influences of aging stereotypes on memory in adulthood. *Psychology and Aging, 21,* 621-625.

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

Inzlicht, M. & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11,* 365-371.

Inzlicht, M. & Ben-Zeev, T. (2003). Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing. *Journal of Educational Psychology, 95,* 796-805.

Jarvis, W. B. G. (2004). MediaLab [Computer Software]. Columbus, OH: Empirisoft.

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

Josephs, R. A., Newman, M. L., Brown, R. P., & Beer, J. M. (2003). Status, testosterone, and human intellectual performance: Stereotype threat as status concern. *Psychological Science, 14,* 158-163.

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

Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting effect on women's math performance. *Personality and Social Psychology Bulletin, 29,* 371-381.

Kemmelmeier, M. & Oyserman, D. (2001). The ups and downs of thinking about a successful other: self-construals and the consequences of social comparisons. *European Journal of Social Psychology, 31,* 310-320.

Leyens, J., Désert, M., Croizet, J-C., & Darcis, C. (2000). Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat? *Personality and Social Psychology Bulletin, 26,* 1189-1199.

Linville, P. W. (1987). Self-complexity as a cognitive buffer against stress-related illness and depression. *Journal of Personality and Social Psychology, 52,* 663-676.

Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin, 28,* 1183-1193.

Mayer, D. M. & Hanges, P. J. (2003). Understanding the stereotype threat effects with "culture-free" tests" An examination of its mediators and measurement. *Human Performance, 16,* 207-230.

McCrea, S. M., & Hirt, E. R. (2001). The role of ability judgments in self-handicapping. *Personality and Social Psychology Bulletin, 27,* 1378-1389.

McIntyre, R. B., Lord, C. G., Gresky, D. M., Ten Eyck, L. L., Frye, G. D. J., & Bond, C. F. (2005). A social impact trend in the effects of role models on alleviating women's mathematics stereotype threat. *Current Research in Social Psychology, 10(9).* Retrieved November 12, 2005, from *http://www.uiowa.edu/%7Egrpproc/crisp/crisp.10.9.html*

McIntyre, R. B., Paulson, R. M., & Lord, C. G. (2003). Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology, 39,* 83-90.

McFarland, L. A., Lev-Arey, D. M., Ziegert, J. C. (2003). An examination of stereotype threat in a motivational context. *Human Performance, 16,* 181-205.

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

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

Ogbu, J. U. (1994). Racial stratification and education in the United States: Why inequality persists. *Teachers College Record, 96,* 264-298.

Oswald, D. L., & Harvey, R. D. (2000). Hostile environments, stereotype threat, and math performance among undergraduate women. *Current Psychology, 19,* 338-355.

Quinn, D. M. & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues, 57,* 55-71.

Sanna, L. J. (1992). Self-efficacy theory: Implications for social facilitation and loafing. *Journal of Personality and Social Psychology, 62,* 774-786.

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

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., & 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.

Schmader., T., Major, B., & Gramzow, R. H. (2001). Coping with ethnic stereotypes in the academic domain: perceived injustice and psychological disengagement. *Journal of Social Issues, 57,* 93-11.

Sekaquaptewa, D. & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's math performance. *Journal of Experimental Social Psychology, 39,* 68-74.

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

Smith, C. E. & Hopkins, R. (2004). Mitigating the impact of stereotypes on academic performance: The effects of cultural identity and attributions for success among

African American college students. *The Western Journal of Black Studies, 28,* 312-321.

Smith, J. L. & White, P. H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement, 61,* 1040-1057.

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

Spencer, S. J., Steele, C. M. & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental and Social Psychology, 35,* 4-28.

Stangor, C., Carr, C., & Kiang, L. (1998). Activating stereotypes undermines task performance expectations. *Journal of Personality and Social Psychology, 75,* 1191-1197.

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

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.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image. In L. Berkowitz (Ed.) *Advances in experimental social psychology.* (pp. 379-440). San Diego, CA: Academic Press.

Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on black and white athletic performance. *Journal of Personality and Social Psychology, 77,* 1213-1227.

Tesser, A. & Campbell, J. (1980). Social definition: The impact of relative importance and similarity of others. *Social Psychology Quarterly, 43*, 341-347.

The College Board. (2003). SAT verbal and math scores up significantly as a record-breaking number of students take the test. Retrieved August 12, 2006 from http://www.collegeboard.com/prof/counselors/tests/sat/resources/resources.html.

Tormala, Z. L., Clarkson, J. J., & Petty, R. E. (2006). Resisting persuasion by the skin of one's teeth: The hidden success of resisted persuasive messages. *Journal of Personality and Social Psychology, 91,* 423-435.

Walsh, M., Hickey, C., & Duffy, J. (1999). Influence of item content and stereotype situation on gender differences in mathematical problem solving. *Sex Roles: A Journal of Research, 41,* 219-240.

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

Wheeler, S. C. & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin, 127,* 797-826.

Zajonc, R. B. (1965). Social facilitation. *Science, 149,* 269-274.

VITA

Personal Background	Dana Prestwood Gresky Fort Worth, Texas Daughter of Alan and Katie Prestwood Married to Aaron Gresky, April 1, 2005
Education	Diploma, Palm Bay High School, Melbourne, Florida, 1997 Bachelor of Science, Psychology, Palm Beach Atlantic College, West Palm Beach, Florida, 2000 Master of Science, Psychology, Texas Christian University, Fort, Worth, Texas, 2003 Doctor of Philosophy, Psychology, Texas Christian University, Fort Worth, Texas, 2006
Experience	Teaching Assistantship, Texas Christian University, Fort Worth, 2001-2005 Adjunct Professor of Psychology, Texas Wesleyan University, Fort Worth, Texas, 2004-2006 Adjunct Professor of Psychology, Tarrant County Community College, Fort Worth, Texas, 2005-2006 Assistant Professor of Psychology, Tenure-track, Texas Wesleyan University, 2006-present
Professional Memberships	Society for Personality and Social Psychology Southwestern Psychological Association Association for Psychological Science

ABSTRACT

THAT WAS HARD! EXAMINING THE EFFECTS OF
TEST INSTRUCTIONS AND CONTENT ON WOMEN'S
MATHEMATICS PERFORMANCE UNDER STEREOTYPE THREAT

by Dana Prestwood Gresky, Ph.D., 2006
Department of Psychology
Texas Christian University

Dissertation Advisor: Charles G. Lord, Professor of Psychology

Previous research on social facilitation suggested that women's mathematics stereotype threat might be alleviated by components of the test itself (Bond, 1982). To test this hypothesis, in Experiment 1, we primed the stereotype about women's math performance, and then gave men and women a difficult math test. In one condition, the format of the math test was intended to lead participants to believe that their performance was good, and in the other condition, the format of the math test was intended to lead participants to believe their performance was poor. Stereotype threatened women who believed they performed poorly on the test actually performed no worse than stereotype threatened women who thought they performed well. In Experiment 2, we primed the stereotype about women's poor math performance, and then gave the men and women a math test. In one condition, the majority of the math items were relatively easy, with some difficult items embedded. In a second condition, the majority of the math items were difficult, including items that matched those used in the first condition. We found that stereotype threatened women performed better on the matched items embedded in an easy test, than did stereotype threatened women completing the same items embedded in a difficult test. Theoretical and practical implications of the results are discussed.