

Gender, Stereotype Threat, and Anxiety: Psychophysiological and cognitive evidence

Jason W. Osborne

Dept. of Curriculum and Instruction, North Carolina State University, Raleigh,
North Carolina, USA

USA

jason_osborne@ncsu.edu

Abstract

Introduction. Claude Steele's stereotype threat hypothesis proposed that negative group stereotypes increase individual anxiety levels, hurting performance. However, the role of anxiety in stereotype threat has not been fully explored. This study examined the hypothesis that experimental manipulation of stereotype threat would influence real-time measures of physiological arousal and cognitive efficiency in girls and boys taking mathematics tests.

Method. Participants were students at a large public university in the USA. Girls and boys were randomly assigned to either high or low stereotype threat conditions, and following an adaptation period, were presented a challenging mathematics task while physiological measures were recorded. Cognitive processing time was recorded for each test item.

Results. Results showed significant physiological reactance (skin conductance, skin temperature, blood pressure) as a function of a stereotype threat manipulation. Results also showed significant differences in cognitive efficiency as a function of stereotype threat.

Conclusion. These findings are consistent with the argument that stereotype threat manipulations either increase or decrease situationally-specific anxiety. These findings hold significant implications for high-stakes academic testing and other situations.

Keywords: gender, stereotype threat, academic performance, anxiety,

Introduction

In the United States, students from disadvantaged minority groups tend to score lower on important academic tasks than Caucasian or Asian students. Students from disadvantaged minority groups tend to receive lower grades in school (Demo & Parker, 1987; Simmons, Brown, Bush, & Blyth, 1978); score lower on standardized tests of intellectual ability (Bachman, 1970; Herring, 1989; Reyes & Stanic, 1988; Simmons et al., 1978), and graduate from college with substantially lower grades than Caucasian students (Nettles, 1988). Ogbu and others have pointed out similar trends for other disadvantaged groups (e.g., Ogbu, 1978; Whitworth & Barrientos, 1990).

Decades of research have attributed performance gaps to factors such as socioeconomic status, academic preparation, and educational opportunities. Yet when background factors are held constant, subsequent achievement is lower for minority students than Caucasian or Asian students (Jensen, 1980; Ramist, Lewis, & McCamley-Jenkins, 1994). Further, achievement gaps are not static, nor do they tend to be present at the beginning of schooling, yet by the sixth year of school there are substantial gaps between Caucasian students and students of color in the US (Alexander & Entwistle, 1988; Valencia, 1991, 1997). Further, data from the Third International Mathematics and Science Study (e.g., Mullis, Martin, Fierros, Goldberg, & Stempler, 2000) shows (a) an increasing performance gap in mathematics and science (particularly physical/mathematical sciences) as students get older, (b) that this effect is pervasive across many countries, but also highly variable across countries, and (c) this pattern holds even for high-performing students and when these results are investigated at the individual item level.

Research has also shown gender disparities in mathematics, sciences, engineering, and technology. While most girls perform as well as boys in general schoolwork in these areas, particularly at earlier ages (Hyde, Fennema, & Lamon, 1990), disparities arise when the material is more advanced and as girls move into high school and college, the career-choosing years (Armstrong, 1981; Benbow & Stanley, 1980, 1983; Ethington & Wolfe, 1984; Fennema & Sherman, 1977, 1978). Women are much less likely to enroll in majors that they perceive to be highly math-focused (LeFevre, Kulak, & Heymans, 1992), even when they score very high on standardized math achievement tests. For example, Turner and Bowen (1999) reported that boys scoring “very high” on mathematics standardized tests are over-represented in

engineering, math, and the physical sciences, whereas girls scoring “very high” on the same tests are over-represented in the biological and non-quantitative social sciences. Further, they conclude that scores on these tests (indicating, presumably, prior preparation in the field of mathematics) accounts for less than half the variance in choice of major.

Stereotype threat theory suggests that that girls may withdraw from, or avoid these classes and majors because math-related anxiety can make them aversive, despite high performance (e.g., Rounds & Hendel, 1980; Spencer, Steele, & Quinn, 1999; Tobias & Weissbrod, 1980).

There are many possible reasons for the aforementioned phenomena, and have often included prior preparation, biological / brain/ genetic differences, and social, or psychological factors (e.g., Benbow & Stanley, 1980, 1983; Eccles, 1987; Levine & Ornstein, 1983). However, these explanations cannot account for the effects demonstrated in the stereotype threat literature.

Stereotype Threat

Claude Steele (1992; 1997) argued that these performance gaps are partly attributable to negative stereotypes concerning group members’ performance. Because of these stereotypes, group members tend to experience higher anxiety on tasks in the stigmatized domain than others not subject to these negative stereotypes. This anxiety is due to the constant fear of being viewed through the lens of the stereotype, of constantly having to fight against being stereotyped, and to worry that any personal failure will be a confirmation of the negative group stereotype.

Increased anxiety not only decreases performance on the task at hand, but also makes the situation aversive to the student leading students to seek escape from the situation either physically by absenteeism or withdrawal, or psychologically via disidentification (for further elaboration on this argument, see Osborne, 1995, 1997; Steele, 1992, 1997). Thus, Steele’s theory may help us understand the differential performance of girls and boys on high-stakes mathematics tests as well as the propensity for girls to shun math-intensive classes and majors in college, despite equal or superior preparation.

Empirical support for Stereotype Threat

There is support for many aspects of Steele's stereotype threat hypothesis. Psychological theory and research supports the basic assumption that awareness of a negative stereotype increases situational anxiety and evaluation apprehension as the domain becomes more self-relevant (Goffman, 1963; Howard & Hammond, 1985; Steele & Aronson, 1995). Further, increased anxiety or arousal can inhibit performance, particularly when that task at hand is complicated or not automatized (Geen, 1991; Hunt & Hillery, 1973; Michaels, Blommel, Brocato, Linkous, & Rowe, 1982; Sarason, 1972; Wigfield & Eccles, 1989) through decreased cognitive capacity, reticence to respond, attentional deficits, and distracting or intrusive thoughts (Geen, 1991; Sarason, 1972). Clawson, Firment, and Trower (1981) observed this effect when they reported that secondary-school students who report higher anxiety tend to score lower on achievement tests in general. Given these findings, two students equal in ability and preparation could show a significant performance disparity if one were to experience stereotype threat while the other did not. The achievement gaps might therefore be partially attributable to the effects of increased anxiety.

Stereotype threat and test performance. Since Steele's hypothesis posits a situationally-specific cause of underperformance, reducing stereotype threat should close the achievement gaps (all other things being equal). The results of experimental investigations into this phenomenon have been encouraging (e.g., Steele & Aronson, 1995). Although much of the discussion and evidence pertains to the gap between African-American and Caucasian students, studies have manipulated stereotype threat in Latino students (Aronson & Salinas, 1997), and girls and women in math and science (Shih, Pittinsky, & Ambady, 1999; Spencer et al., 1999).

Varying the perceived applicability of the stereotype has reduced the gender achievement gap in experimental situations, a finding that is difficult to explain via biological or social phenomena. Spencer, Steele, and Quinn (1999) reported that when a stereotype was perceived to be unrelated to a task the gender gap was substantially reduced compared to when the applicability of the stereotype was not undermined (see also Broadnax, Crocker, & Spencer, 1997). Even highly math-proficient males can experience stereotype threat while taking a math exam. When presenting students with a math test, Aronson, Lustina, Good, Keough, Steele, and Brown (1999) told the participants that the purpose of the experiment was to un-

derstand why Caucasian students did so poorly on the (particular) exam compared to Asian students. As expected, Caucasian males in the stereotype threat condition performed significantly worse than Caucasian males in a no-threat condition.

An intriguing study by of highly math-talented Asian-American female undergraduates by Shih, Pittinsky, and Ambady (1999) demonstrated that Asian-American females' performance on a math achievement test was enhanced when their Asian identity (and hence the positive Asian and math stereotype) was made most salient, and undermined when their female identity (and hence the negative female and math stereotype). Importantly, in this study, groups did not differ on observed motivation, perceptions of test performance, and were not aware that a particular target identity was being made more salient.

These and other studies support the following assertions: (a) stereotype threat is situationally-specific and not a trait of a group, (b) stereotype threat is a phenomenon individuals can experience if they are in a situation where there is a salient negative group stereotype concerning their performance in that domain and the domain is self-relevant, (c) experiencing stereotype threat is aversive, as subjects in these conditions show evidence of escape attempts, (d) acceptance of, or belief in the stereotype is not a necessary condition, and (e) that reducing stereotype threat improves the performance of members of the stigmatized group to the point where performance is often not substantially different from that of non-stigmatized groups once background differences are controlled for. This last point, repeatedly demonstrated in the studies mentioned above, is the main reason why this theory is the focus of much interest and attention.

Evidence for anxiety as the explanatory mechanism in stereotype threat. There are several possible explanations for the observed results. Steele and colleagues (Aronson, Quinn, & Spencer, 1998; Steele, 1997) argue that anxiety explains (mediates) the observed experimental effects summarized above. However, there are other possible explanations.

It could be that a person holds lower expectations or experiences reduced efficacy when the target of a negative group stereotype. A study by Spencer, Steele, and Quinn (1999) tested three possible mediators (self-reported state anxiety, evaluation apprehension, and self-efficacy) in a sample of women and men taking a difficult version of a standardized math test. Results showed that only self-reported anxiety was found to be a partial mediator of the rela-

tionship. Further, the results from Shih et al. (1999) help to rule significant differences in motivation, perceived performance, liking for the test, assessment of test difficulty, or assessment of personal ability.

While much of the research cited above used highly successful college and university students as subjects (making these effects more striking given the restricted range in this population as opposed to the general public-school K-12 population in the United States), Osborne (2001) found that anxiety explained between 38.8% and 41.4% of the racial gap in achievement test scores in a nationally representative sample of high school seniors. Despite the promising results, all of these studies have used self-reported anxiety, raising concerns regarding interpretation and causality.

Only Blascovich, Spencer, Quinn, and Steele (2001) have attempted to measure the hypothesized mediator, anxiety, directly. In this study, Blascovich et al. assessed mean arterial pressure (MAP) reactivity of African-American and Caucasian university students in either low or high stereotype threat conditions. This study showed that African-Americans under the high-stereotype threat condition demonstrated significantly greater MAP reactivity than the other three groups. This study provides direct evidence of physiological reactivity of a particular type (that could be interpreted as indicative of anxiety) while experiencing stereotype threat conditions.

Stereotype Threat and Cognitive Efficiency

Eysenck and Calvo's (1992) Processing Efficiency Theory suggests that as stress or anxiety increases, cognitive efficiency should suffer. Specifically, they argue that anxiety increases task-irrelevant intrusive thoughts that can disrupt the working memory resources and the efficiency of the cognitive process. As cognitive efficiency drops, performance should become worse, or good performance should take longer. This effect should be particularly pronounced when tasks are challenging or performed under a high cognitive load (see also Baddeley & Hitch, 1974; Derakshan & Eysenck, 1998; Hopko, Ashcraft, Gute, Ruggerio, & Lewis, 1998; Klein & Boals, 2001).

Similarly, Hasher and Zacks' (1988) proposed that increasing anxiety leads to more difficulty regulating attention. According to this perspective, while all individuals experience

task-irrelevant or distracting thoughts, more anxious individuals may give these thoughts more attention than less-anxious individuals, also either harming performance or increasing time to perform well. Both perspectives, as well as others (e.g., Geen, 1991; Sarason, 1972), suggest that anxiety can inhibit performance on academic tasks, particularly when those tasks are challenging (i.e., not automatized or overlearned). Following this, individuals experiencing stereotype threat should not only show evidence of increased physiological arousal, but also cognitive sequelae such as increased time to successfully complete tasks or decreased performance on time-limited tasks.

A recent set of studies by Schmader and Johns (2003, Experiment 1) reported that women laboring under a stereotype threat condition showed substantially reduced short-term memory capacity relative to women in a low-stereotype threat condition and men. This research, plus other research on the cognitive and performance effects of anxiety, suggests that stereotype threat manipulations, if truly manipulating anxiety, should not only produce physiological reactance, but also measurable differences in latencies on test items. If memory or other cognitive functioning is at least partially impaired, then individuals laboring under high stereotype threat conditions should take substantially longer to successfully respond to test items than others not laboring under these conditions.

The current study

The literature suggests that Steele's stereotype threat hypothesis might at least partially explain the achievement gaps. While authors have shown several different ways to manipulate stereotype threat (although there are no studies that actually measure stereotype threat, validating that these manipulations are actually affecting only stereotype threat), and significant effects of these manipulations, the mechanism through which stereotype threat works remains to be explicated. While studies like that of Osborne (2001) and Blascovich et al. (2001) suggest the viability of the hypothesis that anxiety is the mediator, neither are ideal. Studies examining anxiety using self-report measures completed following academic activities raise obvious issues of causality and interpretation.

The Blascovich et al. (2001) article is mostly focused on explaining the prevalence of hypertension in the African-American population, and it is open to debate as to whether MAP

reactivity can be interpreted as an indicator of anxiety or not, given the physiological mechanisms controlling MAP (e.g., Brownly, Hurwitz, & Schneiderman, 2000).

The goal of this study was to explore the link between stereotype threat and physiological indicators of anxiety or arousal.

Hypotheses

All students should show signs of increased arousal while taking an academic test if they have any psychological investment in that task. Thus, all hypotheses concerning this study examine change in some variable (e.g., skin conductance) over time that are more dramatic in one group than in another group.

Following Steele's stereotype threat (ST) hypothesis, when girls take a challenging mathematics achievement test under "High ST" conditions (i.e., when the stereotype of female inferiority in mathematics is salient, and girls feel relatively disadvantaged in that domain), anxiety or stress should increase more dramatically than either girls taking the same test under "Low ST" conditions (i.e., when the stereotype of female inferiority on this task is specifically debunked), or boys under either condition. Given this, the first prediction is that girls in the High ST condition should show significantly worse cognitive efficiency than all other groups.

Physiological reactance should similarly be more dramatic for girls in the High ST condition than either girls under Low ST condition or boys. Bradley (2000) gives an excellent overview of the three physiological indicators we are examining, and their underlying neurological, chemical, and physiological mechanisms. The following sections draw heavily from that reference.

Heart rate. According to Bradley (2000), heart rate (HR) is affected by both the sympathetic and parasympathetic nervous systems. Unpleasant visual stimuli tend to produce significant initial HR deceleration, while pleasant or erotic imagery tend to produce initial accelerations. However, HR can be affected by physical fitness, cardiovascular health, hydration,

posture, respiration, and the need for the body to maintain homeostasis and continue life-sustaining activities. There also appears to be a difference between the effects of visual and mental imagery or text-prompted emotion. Text-generated fearful imagery has been shown to produce HR increases that sustain longer than a few seconds.

Bradley (2000) concludes that one can expect heart rate increases to the extent that emotional mental activity is occurring. In the case of this study, Steele's theory indicates that when students are laboring under stereotype threat, there should be increased negative mental activity consistent with anxiety, threat appraisal, or stress. Thus, in the context of this study, girls in High ST condition should show greater increases in HR than girls in Low ST or boys in either condition.

Skin Conductance. Skin conductance (SCL) has been characterized as a pure measure of sympathetic activity, as most of the electrodermal system is controlled exclusively by the sympathetic nervous system (Bradley, 2000; Dawson, Schell, & Filion, 2000). Other authors (e.g., Guyton & Hall, 1996) have argued that palmar sweating, where SCL is usually assessed, might be parasympathetic in nature because it is controlled by a portion of the hypothalamus under control of the parasympathetic nervous system. Regardless, it is clear that SCL measured on the palmar surface of the hands varies dramatically with arousal of either a highly pleasant (e.g., sexual) or highly unpleasant (e.g., violent) nature. SCL changes have also been noted as a function of *anticipation* of pleasant or unpleasant stimuli.

There also appears to be a difference in acclimation to pleasant vs. unpleasant stimuli. Bradley, Kolchakian, Cuthbert, and Lang (1997) showed that reactions to successive positive/pleasant stimuli attenuate over time, whereas reactions to negative/unpleasant stimuli tend to

retain their magnitude (see also Bradley, 2000). Ultimately, if Steele is correct in his hypothesis, SCL should show more dramatic changes in girls testing under High ST conditions than any of the other groups.

Surface skin temperature. Blood vessels tend to constrict when an individual is coping with aversive stimuli, and is most clearly a sympathetic nervous system reaction (Bradley et al., 1997; Brownly et al., 2000). As blood vessels to the skin constrict flow, surface skin temperature will drop moderately, although body and surface skin temperature are slow to change (relative to SCL and HR) and are more specifically bounded by the physiological needs of the body. Life sustaining function simply cannot happen outside a certain narrow temperature range. Thus, girls in High ST conditions should show greater decreases in surface skin temperature (TEMP) at the extremities than other groups, but these changes will be relatively mild and relatively slow compared to other variables.

Method

Participants

Participants were recruited from the psychology pool at a large state university. Participants were compensated with course credit. In all, 42 males and 58 females participated in the study. Participants were randomly assigned to condition through a random number generator, resulting in 50% of males and 50% of the females in each condition. All experimenters were female.

Procedure

Sensor placement. In accordance with recommendations from Dawson et al.(2000), SCL sensors were attached to the volar surfaces of medial phalanges on first two fingers (index, middle) of the non-dominant hand (as students used the dominant hand for working out problems and answering questions). Each sensor cavity was filled sufficiently with biopotential gel created following instructions from Grey and Smith (1984, p. 553). A sensor for heart rate was similarly attached to the volar surface of the medial section of the ring finger on the

same hand, and surface skin temperature was attached to the center of the back of the nondominant hand via surgical tape. The experimenter then verified the equipment was receiving a valid signal from each sensor. Participants were asked not to move their non-dominant hand during the experiment (as excessive movement can cause erroneous readings in the HR monitor).

Acclimation and baseline measurement. Participants spent ten-minutes in an acclimation period reading popular magazines (e.g., Popular Science, Car and Driver, National Geographic, Glamour, Time, etc. Magazines were screened for material of a highly charged nature to prevent baseline measurement error.). After the acclimation period, participants spent the first part of the experiment performing a procedure where they were presented a series of simple photographs and paintings and asked to determine the most common or dominant color in the picture. They responded using the 6-button response pad to get used to using it. During this time the experimenter recorded baseline physiological data.

At the end of the baseline recording period the experimenter verbally administered the experimental intervention.

Experimental Manipulation. Stereotype threat manipulations are generally subtle. Previous research has manipulated: (a) student perceptions as to whether the task assesses academic or intellectual ability or potential (Katz, Roberts, & Robinson, 1965; Steele & Aronson, 1995), (b) perception as to whether the task is diagnostic of ability (Aronson & Tichy, 1997; Quinn & Spencer, 1996), (c) performance prior to testing to emphasize student mastery and improvement in the area (Josephs & Schroeder, 1997), (d) test description to explicitly emphasize or de-emphasize that the task does not show group differences in performance (Broadnax et al., 1997; Spencer et al., 1999) and (e) explicitly describing the test as a measuring of malleable intelligence (Aronson & Fried, 1997; Aronson & Tichy, 1997). In this study stereotype threat was manipulated using the following manipulation (modeled after Spencer et al., 1999), which manipulates perception of whether the tests produce gender differences:

High stereotype threat: As you may know, there has been some controversy about whether there are gender differences in math ability. Previous research has often noted that girls score lower on math tests than boys. We are trying to understand why this might be. We are going to give you two short math tests that most people

find challenging. When students take these tests, girls consistently do worse than boys. You will have seven minutes to complete the first test. We want you to do your best on this test. Take it like you would any other—skip questions you can't answer, feel free to write anywhere on the test. Remember not to move your *[nondominant]* hand while taking the test.

Low stereotype threat: As you may know, there has been some controversy about whether there are gender differences in math ability. Previous research has often noted that girls score lower on math tests than boys. However, there are many cases where girls score as well, or better than boys. We have two short math tests that you will take. Most people find these tests challenging, although these two tests have never shown gender differences. We are trying to understand why this might be. You will have seven minutes to complete the first test. We want you to do your best on this test. Take it like you would any other—skip questions you can't answer, feel free to write anywhere on the test. Remember not to move your *[nondominant]* hand while taking the test.

Participants then began the 20-item mathematics test.

Mathematics achievement test

20 challenging mathematics items from college entrance exams constituted the mathematics achievement test, which was administered via computer and responses were recorded on the response pad. Participants were allowed to spend as much time as they wanted on each item, but were encouraged to work as quickly as possible. Thus, all participants completed all items on this test. Test items were presented in a randomized order to eliminate order effects. This task was designed to be consistently challenging for all participants; in this study, 48% of the items were correctly answered, on average, indicating that our goal was met. Males outscored females with an average of 12.01 to 8.16 items correct ($F_{(1, 87)} = 22.74, p < .0001$). There were no condition effects.

Equipment and physiological measures

Heart rate, skin conductance, and surface skin temperature was recorded using a Biolog 3992 from UFI. The Biolog is a small, battery-powered monitoring unit that is double-shielded from electronic interference with the measurement or recording of the data. The

main unit was attached to sensors with long cords so that it could be placed out of sight of the participant for minimal intrusiveness.

Heart rate. According to Bradley (2000), heart rate (HR) is affected by both the sympathetic and parasympathetic nervous systems. Unpleasant visual stimuli tend to produce significant initial HR deceleration, while pleasant or erotic imagery tend to produce initial accelerations. However, HR can be affected by physical fitness, cardiovascular health, hydration, posture, respiration, and the need for the body to maintain homeostasis and continue life-sustaining activities. There also appears to be a difference between the effects of visual and mental imagery or text-prompted emotion. Text-generated fearful imagery has been shown to produce HR increases that sustain longer than a few seconds.

Bradley (2000) concludes that one can expect heart rate increases to the extent that emotional mental activity is occurring. In the case of this study, Steele's theory indicates that when students are laboring under stereotype threat, there should be increased negative mental activity consistent with anxiety, threat appraisal, or stress. Thus, in the context of this study, girls in High ST condition should show greater increases in HR than girls in Low ST or boys in either condition.

Heart rate was measured through a UFI model 1020 Infrared Pulse Plethysmograph (PPG) transducer that detects heart contractions and ejections by changes in the reflectivity of the skin of the volar surface of the ring finger. It is sampled with 12-bit resolution at 1000Hz to detect the QRS peak. The time between two successive peaks is defined as the interbeat interval, and heart rate calculated as an instantaneous beats per minute (BPM) score from that. The transducer is sensitive to a detection threshold of 0.25 V.

Skin Conductance. Skin conductance (SCL) has been characterized as a pure measure of sympathetic activity, as most of the electrodermal system is controlled exclusively by the sympathetic nervous system (Bradley, 2000; Dawson, Schell, & Filion, 2000). Other authors (e.g., Guyton & Hall, 1996) have argued that palmar sweating, where SCL is usually assessed, might be parasympathetic in nature because it is controlled by a portion of the hypothalamus under control of the parasympathetic nervous system. Regardless, it is clear that SCL measured on the palmar surface of the hands varies dramatically with arousal of either a highly

pleasant (e.g., sexual) or highly unpleasant (e.g., violent) nature. SCL changes have also been noted as a function of *anticipation* of pleasant or unpleasant stimuli.

There also appears to be a difference in acclimation to pleasant vs. unpleasant stimuli. Bradley, Kolchakian, Cuthbert, and Lang (1997) showed that reactions to successive positive/pleasant stimuli attenuate over time, whereas reactions to negative/unpleasant stimuli tend to retain their magnitude (see also Bradley, 2000). Ultimately, if Steele is correct in his hypothesis, SCL should show more dramatic changes in girls testing under High ST conditions than any of the other groups.

Skin conductance was measured via a proprietary UFI voltage excitation SCL signal conditioner that runs a constant 0.5 V across Ag-AgCl electrodes attached to the volar surfaces of the medial phalanges of the subject's nondominant hand. This is the signal and setup recommended by Lykken and Venables (1971), with positioning recommended by Dawson, Schell, & Filion (2000). Biopotential contact medium of the type recommended by Grey and Smith (1984) was used on both electrodes. SCL was sampled at a rate of 10 Hz, with 12-bit resolution, and is sensitive to changes of 0.1 uMho across a range of 0.1 to 40.95 uMho.

Surface skin temperature. Blood vessels tend to constrict when an individual is coping with aversive stimuli, and is most clearly a sympathetic nervous system reaction (Bradley et al., 1997; Brownly et al., 2000). As blood vessels to the skin constrict flow, surface skin temperature will drop moderately, although body and surface skin temperature are slow to change (relative to SCL and HR) and are more specifically bounded by the physiological needs of the body. Life sustaining function simply cannot happen outside a certain narrow temperature range. Thus, girls in High ST conditions should show greater decreases in surface skin temperature (TEMP) at the extremities than other groups, but these changes will be relatively mild and relatively slow compared to other variables.

Surface skin temperature was measured by a UFI model 1070SK solid-state skin temperature transducer. Using surgical tape, the sensor was attached to the center of the back of the non-dominant hand. It is sampled at 1 Hz with 12 bit resolution, and is sensitive to changes of 0.1 degrees Centigrade across a range of 0.1 to 409.6 degrees Kelvin.

Other measures

Perceived performance was assessed at the end of the study via a simple question asking “How do you think you did on these tests?” Responses were gathered on a scale from 1 (very poorly) to 6 (very well).

Manipulation checks assessed at the end of the study asked two questions: “In general, how do you think girls do in math?” and “How do you think girls do on the math tests you took today?” Both items were assessed on a scale from 1 (much worse than boys) to 5 (much better than boys). There were no significant differences in the first question (as expected). For the second question, there was a near-significant trend toward individuals in the High ST condition reporting that girls tend to do less well (mean=2.77) than in the Low ST condition (mean=3.01, $F_{(1,67)} = 2.95, p < .09$).

Data processing

All physiological measures (HR, SCL, TEMP) were measured between one and ten times per second. Data from these channels were aggregated to 30 second intervals for HR and SCL by averaging all valid measurementsⁱ within each 30-second interval, and TEMP was aggregated to 60-second intervals due to the fact that this channel was sampled only once per second.

Hand movements can produce erroneous readings in HR. Therefore, HR measurements were screened for values substantially outside the individual’s range. Readings that had the characteristics of being substantially outside this range (more than double or less than half the preceding values) for only a very brief (less than 3 measurements) period were assumed to be the result of hand movement. These infrequent scores were replaced by the 30-second moving average. Neither of the other two channels suffered this propensity toward artifacts, and although the data were checked for anomalies, none were detected.

Results

Manipulation checks

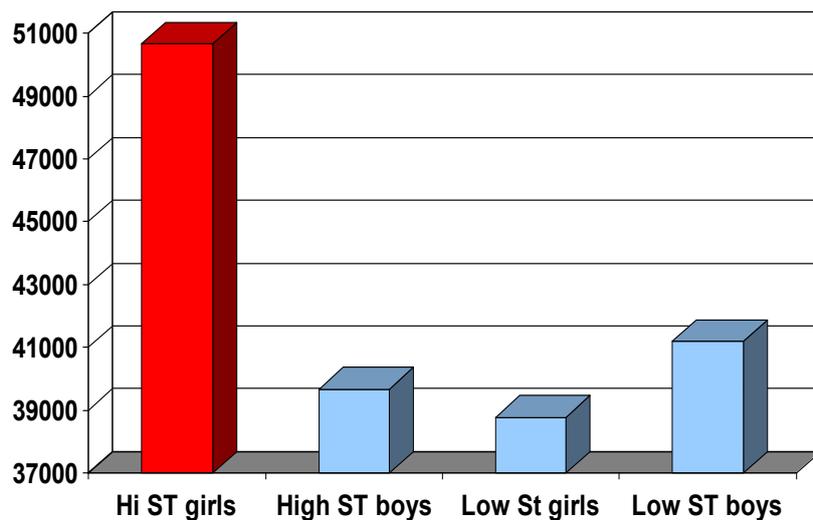
Perceived test performance correlated significantly with actual performance (percent correct, $r = .51, p < .001$), indicating that participants were generally perceiving their performance relatively realistically. Further, as one might expect, test performance and cognitive efficiency

(latencies) were significantly related in a curvilinear relationship indicating that as latencies increased, performance increased to a point, and then decreased with longer latencies ($R = .42$, $F_{(1, 50)} = 5.35$, $p < .008$, prediction equation $y' = -1.17 + 0.0004361 * \text{latency} - 0.000000003941 * \text{latency}^2$ where latency ranged from 17023 to 103735 milliseconds with a median of 42239 milliseconds).

Cognitive efficiency

As all items were presented in a different randomized order for each participant, overall score and time spent on all test items were averaged to form indices of performance and efficiency (the average number of seconds spent on each item). Taking longer to correctly answer an item indicates poorer cognitive efficiency, and is also an indicator of anxiety (Eysenck & Calvo, 1992).

Figure 1.
Response latencies in milliseconds as a function of sex and stereotype threat.



A univariate Analysis of Variance with condition and sex as independent variables and test latency (in milliseconds) as the dependent variable yielded only a significant sex by condition interaction ($F_{(1,50)} = 3.23$, $p < .04$, $\eta^2 = .08$). This analysis was followed by an *a priori* (planned) contrast comparing girls in the High ST- girls condition to the rest of the participants. This contrast was significant ($F_{(150)} = 6.66$, $p < .01$, $\eta^2 = .12$). As presented in Figure 1, girls in the High ST condition had significantly higher latencies than the other groups. Girls who were in the High ST- condition took over 30% longer to answer each individual test item than girls in the low ST condition, and approximately 28% longer than either of the boys groups.

These results are congruent with the results reported by Schmader and Johns (2003), and support the argument that individuals testing under high stereotype threat conditions experience significant amounts of cognitive interference.

Physiological measures

Statistical analyses for these data proceeded from an initial mixed between-within repeated measures analysis of variance, with scores over time as the within-subjects variable and sex and condition as between-subjects factors.

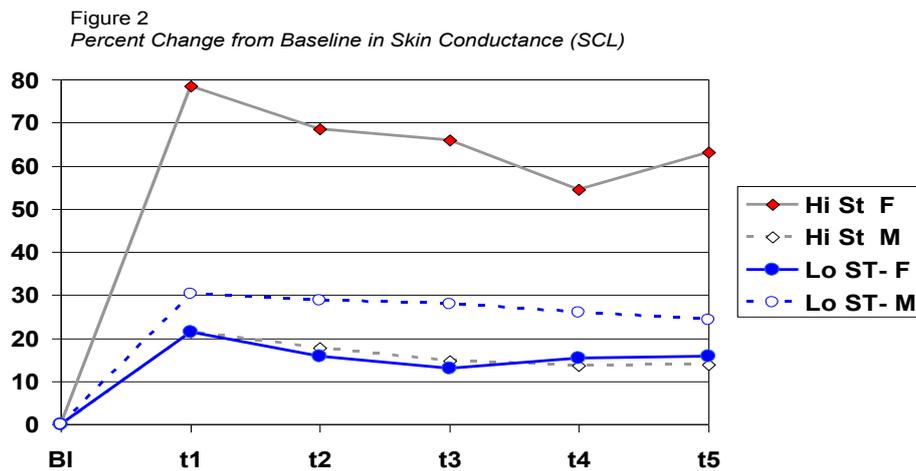
Baseline data. Data for the baseline observation period showed no significant differences as a function of condition in raw SCL (all $p > .15$), HR (all $p > .57$ except sex, which was $p < .11$), and TEMP (all $p > .51$ except sex, which was $p < .09$).

As Dawson, Schell, and Filion (2000, p. 209) note, it is common to adjust skin conductance scores for the range of the individual's baseline skin conductance because SCL and related measures can have large individual differences that is thought to be primarily due to physiological differences in the anatomy of the skin (e.g., thickness of the corneum; similar arguments are possible for other measures with strong, anatomically-driven individual differences, such as heart rate, but not relatively constrained measures such as body temperature), as well as individual differences in hydration at the time of testing (for the seminal discussion of these issues, see also Lykken & Venables, 1971). Therefore, raw SCL scores are generally not of specific interest, but rather, individual variation within that particular individual's range that is of interest in psychological research. Ben-Shakhar (1985) clarified this further in recommending the use of within-subject standardized scores as this relies on a more reliable statistic, a mean score.

Following this recommendation, then, we computed standardized scores to reflect departure from the averaged baseline measurements for SCL and HR, and both measures were converted to percent change from baseline to account for the fact that each individual has a different baseline, and that an individual's baseline can influence the importance of a particular magnitude of change. For example, a SCL reading of 5.0 uMho that changes to 10.0 uMHO is a 100% increase, whereas a 5-point increase from 15.0 uMho to 20.0 uMho, is only a 33% increase, and may reflect a less substantial reaction. The average of the last four minutes of baseline recording for each channel was defined as baseline for that channel for

that individual. Each score after that was converted to a percent increase or decrease from that score $((\text{score}-\text{baseline}) \cdot 100 / \text{baseline})$. This adjusted score reflects the magnitude of deviation from that individual's unique baseline.

Skin conductance. SCL was subjected to a repeated measures ANOVA, with SCL over time as the dependent variable and sex and condition as the independent variables. There was a significant main effect of SCL ($F_{(5, 320)} = 28.47, p < .0001, \eta^2 = .31$), and a significant between-subjects interaction between sex and condition ($F_{(1,50)} = 5.93, p < .004, \eta^2 = .05$) indicating that girls in the High ST condition had higher average SCL than girls in the Low ST conditions (means= 55.14, and 13.50, respectively) or boys in either condition (means=13.57, and 22.93, respectively). As expected, there was also a significant SCL by sex by condition interaction ($F_{(5, 250)} = 2.91, p < .002, \eta^2 = .08$), depicted in Figure 2. No other effect was significant.

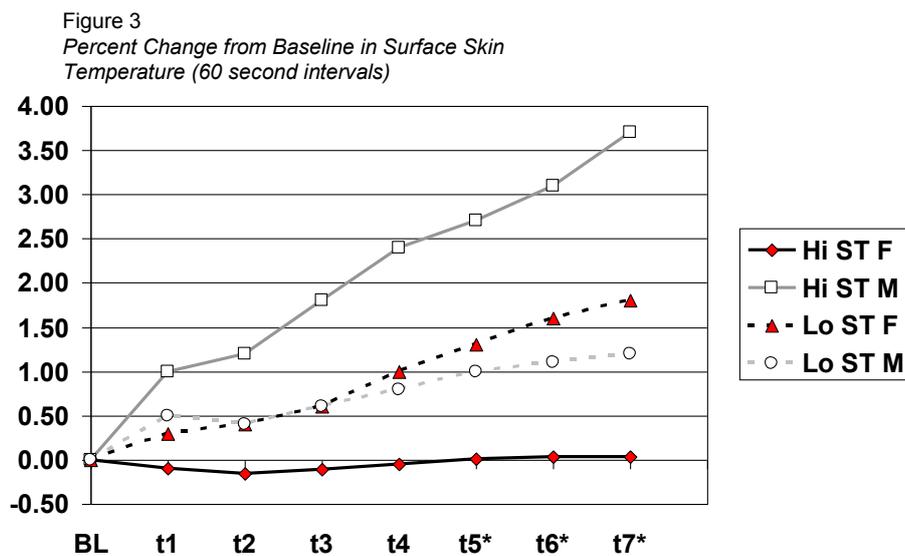


This interaction was explored via univariate *a priori* contrasts comparing girls in the High ST condition with all other groups. All contrasts were significant at levels ranging from $p < .0001$ to $p < .02$.

Heart rate. The analyses of heart rate yielded no significant effects. Trends were in the predicted direction, however.

Surface skin temperature. A repeated measures ANOVA indicated a significant effect for surface skin temperature. Specifically, there was a significant main effect for change in TEMP over time ($F_{(7, 441)} = 29.58, p < .0001, \eta^2 = .32$), and a significant interaction of condi-

tion and sex ($F_{(1,50)} = 4.17, p < .02, \eta^2 = .12$). The interaction of interest was the TEMP by condition by sex interaction ($F_{(7,350)} = 4.33, p < .0001, \eta^2 = .12$) indicating that girls in the High ST condition showed the expected decrease in surface skin temperature, while other groups showed increases (presented in Figure 3). Post-hoc comparisons revealed significant contrasts between girls in the High ST condition and other groups for the last 3 time periods (marked on the Figure with an asterisk). This is understandable as skin temperature is a relatively slow-changing variable compared to heart rate and SCL.



Discussion

This study was designed to explore Steele's hypothesis that students subjected to stereotype threat might experience relatively inflated levels of anxiety, stress, or arousal, and that this increased physiological arousal should impair cognitive efficiency. Consistent with expectations, the results of this study show evidence of physiological reactivity for girls under high stereotype threat conditions that are consistent with an anxiety or autonomic arousal reaction. Substantial and significant reactivity in skin conductance (SCL), surface skin temperature (TEMP), and diastolic blood pressure (DBP) all support Steele's argument regarding the nature of stereotype threat, and its possible effects on academic performance. These results were also congruent with the findings by Blascovich et al. (2001).

Further, these results show a plausible link between the manipulation of stereotype threat and actual academic performance. As multiple theories (discussed above) point out, increased anxiety should decrease cognitive efficiency, which should thus either cause students to take longer to perform at par with students not laboring under increased anxiety, or, when tasks are time-limited, to perform more poorly, despite being equally capable and prepared. This second finding holds significant implications for academic testing, as many tests (particularly high-stakes tests) tend to be time-limited. It is not surprising that in the US there are significant performance gaps on high-stakes achievement tests when our results show it takes stigmatized groups up to 30% longer per item to answer.

One of the important lessons from this and other stereotype threat studies is that under *low* stereotype threat conditions, achievement gaps tend to attenuate. Given that these manipulations are often quite subtle (i.e., merely telling students that tasks are fair or specifically debunking stereotypes of inferiority), there is reason to hope that subtle changes to testing paradigms can substantially improve the opportunity for students from stigmatized groups to perform to their potential.

Caveats and directions for future research

It is important to remember that while it is intuitive that anxiety and stress are related to physiological reactions, there is controversy and debate in the literature regarding the relationship between physiological changes and changes in emotionality. The safest interpretation of these results is that physiological reactance was increased under high stereotype threat conditions, relative to low stereotype threat conditions. Labeling these reactions as indicative of anxiety is not technically supported by the science of psychophysiology at this point. Nevertheless, these results are consistent with Steele's stereotype threat hypothesis.

Second, these studies investigated one particular aspect of stereotype threat (sex and math performance). Although it seems intuitive to generalize to other instances (e.g., race, age), that should be done with care.

Furthermore, it is reasonable to wonder whether experimenter sex may have had an influence in the reactions of the participants (following research by Inzlicht & Ben-Zeev, 2000 and others). In this study all experimenters were female, and thus we held sex constant across

condition. However, this in itself represents an important line of research, especially as stereotype threat relate to education. Is it the case, for example, that a female math teacher can reduce stereotype threat more effectively than a male teacher?

One must note that the participants in this study were University students, and therefore represent a select segment of the student population most likely to be successful in academics. However, as many of these participants were psychology majors, many of whom are notoriously math-averse, it may not be as difficult a generalization as one might expect. Additionally, this sample, if it represented only the most successful segment of the student population, would therefore bias the results toward the null hypothesis, rather than advantaging the results toward rejection of the null. Having found some substantial effects within this restricted population, it makes another study using secondary school students an interesting prospect and important direction for future research.

These studies did not replicate earlier findings showing that manipulation of stereotype threat can produce differences in performance on academic tasks. The difference was that participants had unlimited time on each item. It is likely (given the data on latencies showing girls in the high -stereotype threat condition took substantially longer to answer than participants in other conditions) that had this test been time-limited, results would have replicated previous studies.

Finally, as with much of this type of research, within-group variance was substantial, hurting statistical inference. Some individuals in high stereotype threat conditions showed virtually no reaction, while others responded dramatically. Likewise, some individuals in the low stereotype threat condition showed reactance where others did not. This field and this paradigm needs to move to a more individualistic approach to understanding the effects of stereotypes on an individual level. Future research needs to look at what individual differences predict whether an individual will react or be sensitive to stereotypes, and what individual differences might be protective against these effects. Researchers have identified several variables that are related to individual differences in reactance to stereotype threat. These include stereotype relevance (Brown & Josephs, 1999), gender identification (Schmader, 2002), and others, including identification with academics (e.g., Steele, 1997). Identification with academics was assessed in both studies, but unfortunately the college populations participating in the study had a relatively restricted range in identification with academics (all scoring rela-

tively high). Thus, moderation effects could not be examined, and covarying identification with academics failed to produce any alteration in the nature of the results. A more sensitive measure of identification with mathematics performance would have been desirable (as used in Brown & Josephs, 1999), and future research should attempt to measure domain identification as sensitively and narrowly as possible.

However, not having this measure, and thus, not being able to separate out less identified from more identified individuals will only add error variance to the data, decreasing the likelihood of rejecting the null hypothesis. Since significant trends were observed, although unfortunate, this is not a fatal flaw.

Conclusions

In the USA it is virtually compulsory for states to participate in high-stakes testing of students from very early in their academic career. Given the stereotype threat literature, and other theoretical and methodological reasons that contraindicate the use of high-stakes testing in public education, educators are left in an interesting dilemma. If stereotype threat research is to be taken seriously, then one must seriously consider the notion that widely-reported “achievement gaps” between various groups may be little more than the effect of relatively impoverished backgrounds and societal stereotypes. This requires us to carefully question high-stakes testing in education.

On the positive side, stereotype threat research has demonstrated that minimal alterations to testing situations can substantially reduce the observed achievement gaps, at least in research settings. This should be pursued in order to engineer testing and classroom situations to minimize stereotype threat, or maximize student resilience in the face of ubiquitous negative group stereotypes. It has now been almost 15 years since Steele’s (1992) *Atlantic Monthly* article announced his formulation of his Stereotype Threat hypothesis. The literature has demonstrated the importance of the idea. Implementation of the principles and lessons from a decade of research cannot come quickly enough.

References

- Alexander, K. L., & Entwistle, D. R. (1988). Achievement in the first two years of school: Patterns and processes. *Monographs of the Society for Research in Child Development, 53*(2).
- Armstrong, J. M. (1981). Achievement and participation of women in mathematics: Results from two national surveys. *Journal of Research in Mathematics Education, 12*, 356-372.
- Aronson, J., & Fried, C. (1997). *Belief in the malleability of intellectual ability: An intervention to increase the performance of African-American college students*. Unpublished manuscript.
- 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*(1), 29-46.
- Aronson, J., Quinn, D. M., & Spencer, S. J. (1998). Stereotype threat and the academic underperformance of minorities and women. In J. K. Swim & C. Stangor (Eds.), *Prejudice: The Target's Perspective*. New York: Academic Press.
- Aronson, J., & Salinas, M. F. (1997). *Stereotype threat, attributional ambiguity, and Latino underperformance*. Unpublished manuscript.
- Aronson, J., & Tichy, J. C. (1997). *Conceptions of ability and stereotype threat*. Unpublished manuscript.
- Bachman, J. G. (1970). *Youth in Transition: Volume 2. The impact of family background and intelligence on 10th-grade boys*. Ann Arbor, Michigan: Survey Research Center, Institute for Social Research.
- Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York: Academic Press.
- Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science, 210*, 1262-1264.
- Benbow, C. P., & Stanley, J. C. (1983). Sex differences in mathematical reasoning ability: More facts. *Science, 222*, 1029-1031.
- Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science, 12*(3), 225-229.
- Bradley, M. M. (2000). Emotion and Motivation. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of Psychophysiology, 2nd Edition*. Cambridge, UK: Cambridge University Press.

- Bradley, M. M., Kolchakian, M., Cuthbert, B. N., & Lang, P. J. (1997). What's new? What's exciting? Novelty and emotion in perception. *Psychophysiology*, *34*, 1-22.
- Broadnax, S., Crocker, J., & Spencer, S. J. (1997). *African-Americans and academic performance: The mediational role of stereotype vulnerability*. Unpublished manuscript.
- 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*(2), 246-257.
- Brownly, K. A., Hurwitz, B. E., & Schneiderman, N. (2000). Cardiovascular psychophysiology. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of Psychophysiology*, 2nd Edition. Cambridge, UK: Cambridge University Press.
- Clawson, T. W., Firment, C. K., & Trower, T. L. (1981). Test anxiety: Another origin for racial bias in standardized testing. *Measurement and Evaluation in Guidance*, *13*, 210-215.
- Dawson, M. E., Schell, A. M., & Filion, D. L. (2000). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of Psychophysiology*, 2nd Edition. Cambridge, UK: Cambridge University Press.
- Demo, D. H., & Parker, K. D. (1987). Academic achievement and self-esteem among African-American and White college students. *Journal of Social Psychology*, *127*(4), 345-355.
- Derakshan, N., & Eysenck, M. W. (1998). Working memory capacity in high trait-anxious and repressor groups. *Cognition and emotion*, *12*(5), 697-713.
- Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly*, *11*, 135-172.
- Ethington, C. A., & Wolfe, L. M. (1984). Sex differences in a causal model of mathematics achievement. *Journal for Research in Mathematics Education*, *15*, 361-377.
- Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and emotion*, *6*, 409-434.
- Fennema, E. H., & Sherman, J. A. (1977). Sex-related differences in mathematics achievement, spatial visualization, and sociocultural factors. *American Educational Research Journal*, *14*(51-71).
- Fennema, E. H., & Sherman, J. A. (1978). Sex-related differences in mathematics achievement: A further study. *Journal for Research in Mathematics Education*, *9*, 189-203.
- Geen, R. G. (1991). Social Motivation. *Annual Review of Psychology*, *42*, 377-399.

- Goffman, E. (1963). *Stigma: Notes on the management of a spoiled identity*. New York: Touchstone.
- Grey, S. J., & Smith, B. L. (1984). A comparison between commercially available electrode gels and purpose-made gel in the measurement of electrodermal activity. *Psychophysiology*, *21*, 551-557.
- Guyton, A. C., & Hall, J. E. (1996). *Textbook of Medical Physiology*. Philadelphia, PA.: Saunders.
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *The Psychology of Learning and Motivation*, *22*(193-225).
- Herring, R. D. (1989). Minimum competency testing: Implications for ethnicity. *Journal of Multicultural Counseling and Development*, *17*, 98-104.
- Hopko, D. R., Ashcraft, M. H., Gute, J., Ruggerio, K. J., & Lewis, C. (1998). Mathematics anxiety and working memory: Support for the existence of a deficient inhibition mechanism. *Journal of Anxiety Disorders*, *12*(4), 343-355.
- Howard, J., & Hammond, R. (1985, September 9). Rumors of inferiority. *The New Republic*, *193*, 17-21.
- Hunt, P. J., & Hillery, J. M. (1973). Social facilitation in a location setting: An examination of the effects over learning trials. *Journal of Experimental Social Psychology*, *30*, 139-155.
- 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.
- Jensen, A. R. (1980). *Bias in Mental Testing*. New York: Free Press.
- Josephs, R., & Schroeder, D. (1997). *The self-protective function of the learning curve*. Unpublished manuscript.
- Katz, I., Roberts, S. O., & Robinson, J. M. (1965). Effects of task difficulty, race of administrator, and instructions on digit-symbol performance of Negroes. *Journal of Personality and Social Psychology*, *2*, 53-59.
- Klein, K., & Boals, A. (2001). The relationship of life events stress and working memory capacity. *Applied Cognitive Psychology*, *15*, 565-579.

- LeFevre, J., Kulak, A., & Heymans, S. (1992). Factors influencing the selection of university majors varying in mathematical content. *Canadian Journal of Behavioural Science*, 24, 276-289.
- Levine, D. U., & Ornstein, A. C. (1983). Sex differences in ability and achievement. *Journal of Research and Development in Education*, 16, 62-72.
- Lykken, D. T., & Venables, P. H. (1971). Direct measurement of skin conductance: A proposal for standardization. *Psychophysiology*, 8, 656-672.
- Michaels, J. W., Blommel, J. M., Brocato, R. M., Linkous, R. A., & Rowe, J. S. (1982). Social facilitation and inhibition in a natural setting. *Replications in Social Psychology*, 2(21-24).
- Mullis, I.V., Martin, M. O., Fierros, E. G., Goldberg, A. L., & Stemler, S. E. (2000). *Gender differences in achievement: IEA's Third International Mathematics and Science Study (TIMMS)*. TIMMS International Study Center, Boston College, Chestnut Hill: MA.
- Nettles, M. T. (1988). *Toward undergraduate student equality in American higher education*. New York: Greenwood.
- Ogbu, J. U. (1978). *Minority education and caste: The American system in cross-cultural perspective*. New York: Academic Press.
- Osborne, J. W. (1995). Academics, self-esteem, and race: A look at the assumptions underlying the Disidentification hypothesis. *Personality and Social Psychology Bulletin*, 21(5), 449-455.
- Osborne, J. W. (1997). Race and academic disidentification. *Journal of Educational Psychology*, 89(4), 728-735.
- Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology*, 26(3), 291-310.
- Osborne, J. W. (2004). Identification With Academics and Violence in Schools. *Review of General Psychology*, 8(3), 147-162.
- Quinn, D., & Spencer, S. J. (1996). *Stereotype threat and the effect of test diagnosticity on women's math performance*. Paper presented at the American Psychological Association, Toronto, Canada.
- Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1994). *Student group differences in predicting college grades: Sex language and ethnic groups (College Board Report No. 93-1, ETS No. 94.27)*. New York: College Entrance Examination Board.

- Reyes, L. H., & Stanic, G. M. (1988). Race, sex, socioeconomic status, and mathematics. *Journal for Research in Mathematics Education, 19*, 26-43.
- Rounds, J. B., & Hendel, D. D. (1980). Measurement and dimensionality of mathematics anxiety. *Journal of Counseling Psychology, 37*, 138-149.
- Sarason, I. G. (1972). Experimental approaches to test anxiety: Attention and the uses of information. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research* (Vol. 2). New York, NY: Academic Press.
- 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.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80-83.
- Simmons, R. G., Brown, L., Bush, D. M., & Blyth, D. A. (1978). Self-esteem and achievement of Black and White adolescents. *Social Problems, 26*, 86-96.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4-28.
- Steele, C. M. (1992, April). Race and the schooling of Black Americans. *The Atlantic Monthly, 68*-78.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*(6), 613-629.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797-811.
- Tobias, S., & Weissbrod, C. (1980). Anxiety and mathematics: An update. *Harvard Educational Review, 50*(1), 63-80.
- Turner, S. E., & Bowen, W. G. (1999). Choice of Major: The Changing (Unchanging) Gender Gap Source. *Industrial & labor relations review, 52*(2), 289-313.
- Valencia, R. R. (1991). The plight of Chicano students: An overview of schooling conditions and outcomes. In R. R. Valencia (Ed.), *Chicano school failure and success: Research and policy agendas for the 1990s* (pp. 3-26). London, England: Falmer Press.
- Valencia, R. R. (1997). Latinos and education: An overview of sociodemographic characteristics and schooling conditions and outcomes. In M. Barrera-Yepes (Ed.), *Latino education issues: Conference proceedings*. Princeton, NJ: Educational Testing Service.

- Whitworth, R. H., & Barrientos, G. A. (1990). Comparison of Hispanic and Anglo graduate record examination scores and academic performance. *Journal of Psychoeducational Assessment*, 8(2), 128-132.
- Wigfield, A., & Eccles, J. S. (1989). Test anxiety in elementary and secondary school students. *Educational Psychologist*, 24, 159-183.

¹ In the case of some physiological measurement, there are occasionally missed or dropped measurements due to things such as participant finger movement. These are fairly rare, and in this study were primarily missed heart rate measurements.

[This page intentionally left blank]
