Educational Psychology

ISSN: 0144-3410 (Print) 1469-5820 (Online) Journal homepage: www.tandfonline.com/journals/cedp20

Linking Stereotype Threat and Anxiety

Jason W. Osborne

To cite this article: Jason W. Osborne (2007) Linking Stereotype Threat and Anxiety, Educational Psychology, 27:1, 135-154, DOI: 10.1080/01443410601069929

To link to this article: https://doi.org/10.1080/01443410601069929

Published online: 16 May 2007.

Educational Psychology
Vol. 27, No. 1, February 2007, pp. 135–154

Linking Stereotype Threat and Anxiety

Jason W. Osborne*
North Carolina State University, USA

Claude Steele's stereotype threat hypothesis has attracted significant attention in recent years. This study tested one of the main tenets of his theory—that stereotype threat serves to increase individual anxiety levels, thus hurting performance—using real-time measures of physiological arousal. Subjects were randomly assigned to either high or low stereotype threat conditions involving a challenging mathematics task while physiological measures of arousal were recorded. Results showed significant physiological reactance (skin conductance, skin temperature, blood pressure) as a function of a stereotype threat manipulation. These findings are consistent with the argument that stereotype threat manipulations either increase or decrease situational-specific anxiety, and hold significant implications for thinking about fair assessment and testing practices in academic settings.

Students from disadvantaged minority groups tend to score lower on important academic outcomes than Caucasian or Asian students. In the United States, students of colour tend to receive lower grades in school (Demo & Parker, 1987; Simmons, Brown, Bush, & Blyth, 1978), score lower on standardised 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. By the sixth year of school (in the USA) there are substantial gaps between Caucasian students and students of colour (Alexander & Entwhistle, 1988; Valencia, 2002).

*Department of Curriculum and Instruction, North Carolina State University, Poe Hall 602, Campus Box 7801, Raleigh NC 27695-7801, USA. Email: jason_osborne@ncsu.edu

ISSN 0144-3410 (print)/ISSN 1469-5820 (online)/07/010135–20

DOI 10.1080/01443410601069929

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 mathematics-focused (LeFevre, Kulak, & Heymans, 1992), possibly because maths-related anxiety can influence students to withdraw from, or avoid, classes relating to maths, and can influence occupational aspirations (Rounds & Hendel, 1980; Spencer, Steele, & Quinn, 1999; Tobias & Weissbrod, 1980).

Similar to explanations for racial gaps, explanations of the gender gap have often looked to biological, genetic, social, or psychological factors (Benbow & Stanley, 1980, 1983; Eccles, 1987; Levine & Ornstein, 1983).

Stereotype Threat

Claude Steele (1992, 1997) argued that these performance gaps are at least partly attributable to negative stereotypes concerning group members' performance. Because of these stereotypes, group members tend to experience higher anxiety on tasks in the stigmatised 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, to constantly having to fight against being stereotyped, and to the 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 potentially explains many of the interesting research findings of group differences in achievement testing and other academic outcomes, and also can help us to understand differential dropout rates, and even, perhaps, differential prevalence of antisocial behaviour (Osborne, 2004; Osborne & Rausch, 2001). This perspective may help us to understand the differential performance of girls and boys on high-stakes maths tests, as well as the propensity of girls to shun maths-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 the task at hand is complicated or not automatised (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. 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. Given these findings, two students equal in ability and preparation could show a significant discrepancy if one were to experience stereotype threat while the other did not. The racial and gender 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 maths and science (Shih, Pittinsky, & Ambady, 1999; Spencer et al., 1999).

Varying the perceived applicability of the stereotype has also reduced the gender achievement gap in experimental situations. Spencer et al. (1999) reported that when a stereotype was perceived to be unrelated to a task the gender gap was less than when the applicability of the stereotype was not undermined (see also Broadnax, Crocker, & Spencer, 1997). Even highly maths-proficient males can experience stereotype threat while taking a maths exam. When presenting students with a math test, Aronson et al. (1999) told the participants that the purpose of the experiment was to understand 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 of highly maths-talented Asian-American female undergraduates by Shih et al. (1999) demonstrated that Asian-American females' performance on a maths achievement test was enhanced when their Asian identity (and hence the positive Asian and maths stereotype) was made most salient, but undermined when their female identity (and hence the negative female and math stereotype) was salient. Importantly, in this study the groups did not differ on observed motivation or 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: (1) stereotype threat is situationally-specific and not a trait of a group; (2) 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; (3) experiencing stereotype threat is aversive, as subjects in these conditions show evidence of escape attempts; (4) acceptance of, or belief in, the stereotype is not a necessary condition; and (5) reducing stereotype threat improves the performance of members of the stigmatised group to the point where performance is often not substantially different from that of non-stigmatised 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 summarised 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. The study by Spencer et al. (1999) tested three possible mediators (state anxiety, evaluation apprehension, and self-efficacy) in a sample of women and men taking a difficult version of a standardised maths test. Results showed that only anxiety was found to be a partial mediator of the relationship. Further, the results from Shih et al. (1999) help to rule out 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 primary and secondary school population in the United States), Osborne (2001) found that anxiety explained 38.8–41.4% 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 hypothesised mediator, anxiety, directly. In this study, Blascovich et al. assessed the 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.

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, and significant effects of these manipulations, the mechanism through which stereotype threat works remains to be explicated.[1] While studies like those 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 debatable 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). However, it is the best support for the anxiety and stereotype threat link thus far.

Manipulating Stereotype Threat

Stereotype threat manipulations are generally subtle. Previous research has manipulated: (1) student perceptions as to whether the task assesses academic or intellectual ability or potential (Katz, Roberts, & Robinson, 1965; Steele & Aronson, 1995); (2) perception as to whether the task is diagnostic of ability (Aronson & Tichy, 1997; Quinn & Spencer, 1996); (3) performance prior to testing to emphasise student mastery and improvement in the area (Josephs & Schroeder, 1997); (4) test description to explicitly emphasise or de-emphasise that the task does not show group differences in performance (Broadnax et al., 1997; Spencer et al., 1999); and (5) test description to emphasise that the test measures malleable intelligence (Aronson & Fried, 1997; Aronson & Tichy, 1997).

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 in this study examine change in some variable (e.g., skin conductance) over time that is 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 in girls taking the same test under low ST conditions (i.e., when the stereotype of female inferiority on this task is specifically debunked) or in boys under either condition.

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 tends 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 and text-prompted emotion. Text-generated fearful imagery has been shown to produce HR increases that last longer than a few seconds.

Bradley (2000) concludes that one can expect heart rate increases to correspond to emotional mental activity. In the case of this study, Steele's theory indicates that when students are labouring 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 the high ST condition should show greater increases in HR than girls in the low ST or boys in either condition.

Skin conductance. Skin conductance (SCL) has been characterised 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 the 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 the anticipation of pleasant or unpleasant stimuli.

There also appears to be a difference in acclimatisation 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, blood flow to the skin reduces, and 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 the high ST conditions should show greater decreases in surface skin temperature (TEMP) at their extremities than other groups, but these changes will be relatively mild and relatively slow compared to other variables.

Blood pressure. Blood pressure is affected by many factors, including constriction or relaxation of blood vessels, which is most relevant for our purposes. Following the above hypothesis regarding surface skin temperature and constriction of blood vessels, girls in high ST conditions should show increases in either systolic blood pressure (SBP) or diastolic blood pressure (DBP).

Method

Sample and Procedures

The sample consisted of 43 students recruited from an undergraduate psychology pool at a large state university in exchange for course credit (10 males, 14 females in high ST, 11 males and 8 females in low ST; difference in group sizes reflect subject pool demographics as well as the effects of random assignment).

Prior to meeting the participant, the male experimenter flipped a coin to determine at random the condition the participant would enter. Participants were greeted by the experimenter and seated in a room with a large table and chair. The experimenter asked whether the participant was left- or right-handed. Jewellery and watches were removed from the non-dominant hand (if possible).

Sensor placement. In accordance with recommendations from Dawson et al. (2000), SCL sensors were attached to the volar surfaces of the medial phalanges on the first two fingers (index, middle). 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, and surface skin temperature was monitored via a sensor attached to the centre of the back of the hand via surgical tape. A blood pressure cuff was attached to the dominant arm over the brachial artery (in a deflated state). 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).

Baseline measurement. Participants spent 10 minutes in an acclimatisation 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 acclimatisation period the experimenter asked the participants to sit up straight, with both feet flat on the floor, and took a baseline measurement of blood pressure. Then the experimenter stated, "We need to take a few minutes to record some initial data. When we're done, we will begin the experiment. Relax and try not to move your [nondominant] arm." The experimenter started recording data for a five-minute baseline assessment of physiological data. No data displays were visible to the participants. Participants were free to continue their reading during this time.

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

Experimental Manipulation

In this study stereotype threat was manipulated using the following manipulation (from Spencer et al., 1999), which manipulates perception of whether the tests produce gender differences.

In the high ST condition, students were told:

> 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. When seven minutes are up, I will stop you, you will have a short rest period, and then you will take the second test. Mark your answers right on the test by circling the correct answer.

In the high ST condition, students were told:

> 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. When seven minutes are up, I will stop you, you will have a short rest period, and then you will take the second test. Mark your answers right on the test by circling the correct answer.

The experimenter gave the participants the first test and scratch paper, started a digital stopwatch, and simultaneously marked the start of the first test on the physiological recording device via an event marker in the data stream. When seven minutes were up the experimenter collected the test and measured blood pressure. Another test was administered with the same seven-minute time limit. Following completion of this test subjects were then debriefed according to human subjects guidelines, and released.

Equipment and Physiological Measures

Heart rate, skin conductance, and surface skin temperature were recorded using a Biolog 3992 from UFI. Blood pressure was measured via a separate digital sphygmomanometer, described below. 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 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 0.25V.

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

Skin conductance was measured via a proprietary UFI voltage excitation SCL signal conditioner that runs a constant 0.5V 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 et al. (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 10Hz, with 12-bit resolution, and is sensitive to changes of 0.1uMho across a range of 0.1–40.95uMho.

Blood pressure was measured via a Samsung HD-2000F digital blood pressure gauge. This is an automatic inflating unit that utilises the oscillometric method of measurement and has a noise cancellation tank to enhance sensitivity. Studies have shown this gauge to be not significantly different from a trained nurse using a standard sphygmomanometer. The HD-2000f has a measurement range of 20–320 mmHg and accuracy to within ± 2mmHg.

Mathematics Achievement Test

The goal of this study was not to replicate the many studies that have demonstrated that stereotype threat manipulations can affect achievement or academic test performance. Rather, the goal of this study was to examine whether there was evidence of physiological changes as a function of stereotype threat manipulation. Therefore, these mathematics tasks were constructed to be very challenging.

Two short mathematics tests were constructed from recent GRE (Graduate Record Examinations, common graduate school aptitude tests in the USA) mathematics sections. The first test was modelled on routine multiple-choice items. In constructing this mathematics test we chose 10 items, of which five were relatively easy (> 70% of test takers got the item correct) and five were challenging (< 30% of test takers got the item correct). For the second test, we chose items that required students to indicate if one function was equal to, greater than, or less than a second function, or if it was impossible to determine. For this test we chose 16 items, eight relatively easy and eight challenging, as defined above. Students were given only seven minutes for each test, which meant that most students were significantly time-pressured to complete the test, and most did not.

Other Measures and Manipulation Checks

Identification with academics was measured via 36 items drawn from the School Perceptions Questionnaire (Osborne, 1997), which measures the psychological centrality of a domain to the self concept (strongly influenced by the symbolic interactionist perspective of self esteem), and Voelkl's (1996, 1997) Identification with School scale, which was designed to assess student feelings of belongingness (i.e., acceptance of and respect for the self by others in the school, feelings of inclusion) and valuing (i.e., the extent to which the student views schooling as an important institution in society, the material being learned as important and useful, etc.). Both scales have shown good reliability and validity (see Osborne, 1997; Voelkl, 1997) and combine as a unitary dimension with a Cronbach's alpha of .92 (in this sample). All items were measured on a scale of 1 (strongly disagree) to 5 (strongly agree), and items worded negatively were reverse coded prior to averaging. Missing data were handled through mean substitution. Because identification with academics is theoretically important (e.g., as identification with academics increases, the effect of stereotype threat increases) this variable was covaried in all analyses.

Perceived performance was assessed at the end of the study via an item asking: "How do you think you did on these tests?" Responses were gathered on a scale from 1 (very poorly) to 5 (very well). Responses to this item correlated significantly with actual performance (% correct, $r = .41$, $p < .001$), indicating that participants were generally perceiving their performance relatively realistically. There were no sex or condition differences in this relationship.

After the perceived performance question, subjects were asked to respond to "In general, how do you think girls do in math?" and "How do you think girls do on the math tests you took today?" on a scale from 1 (much worse than boys) to 5 (much better than boys). Analyses of responses to the first item showed no significant effects across condition, sex, or the interaction of the two, as expected (all $p > .22$). Analyses of the second question showed that most subjects answered either 2 (a little worse than boys) or 3 (neither better nor worse than boys); one answered 1 and eight answered 4 (a little better than boys). A factorial ANOVA of this variable (with sex and condition as factors) showed a trend for participants in the high ST condition to score lower (average = 2.72) than those in the low ST condition (average = 3.08, $F_{(1,50)}=3.20$, $p < .08$). Since the manipulation was designed to set up a binary response (worse than boys, not worse than boys) the data were recoded into a binary outcome (1 = worse than boys, 2 = not worse than boys or better than boys) and subjected to logistic regression analysis, with sex, condition, and the interaction as predictors. As expected, this analysis revealed that subjects in the high ST condition were 6.77 times as likely as those in the low ST condition to respond that girls should do worse than boys on the tests ($F_{(1, 50)} = 5.27$, $p < .020$). This represents evidence of the efficacy of the minimal ST manipulation (all assumptions were checked and met for aforementioned analyses).

Data Processing

Heart rate (HR), skin temperature (SCL), and surface skin temperature (TEMP) were measured 1–10 times per second. Data from these channels were aggregated to 30-second intervals for HR and SCL by averaging all valid measurements[2] 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. Blood pressure was only measured at baseline, and after the completion of the first test section, since it required the subject to sit still for a period of time.

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

Results

Test Results

Students answered between 11 and 26 of the 26 items (10 items from the first part of the test, 16 items from the second part of the test), averaging 20.2 with a standard deviation of 4.60. There were no significant differences in the number of items attempted as a function of sex or condition (all $p > .35$). Students gave 3–21 correct answers, averaging 8.3 with a standard deviation of 3.43. As expected and reported in previous studies of this nature, performance was significantly related to sex and condition ($F_{(1,50)} = 5.46$, $p < .02$, $\eta^2 = .12$). As shown in Figure 1, there is a large gap in performance between males and females under high ST conditions (means = 11.28 and 6.87 respectively) and no substantial difference between males and females under low ST conditions (means = 8.70 and 9.47 respectively).

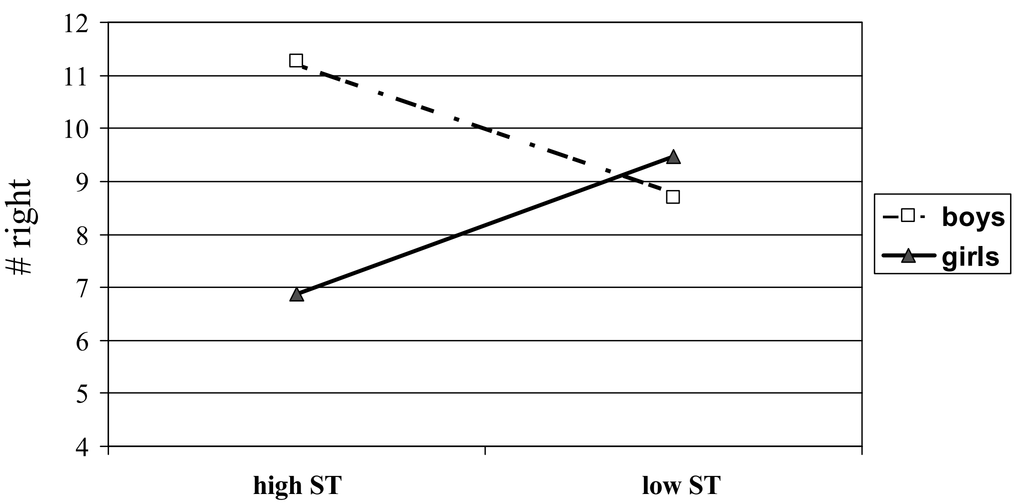

Figure 1. Stereotype threat and test performance

Physiological Indices

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 et al. (2000) note, it is common to adjust skin conductance scores to the range of the individual's baseline skin conductance, because SCL and related measures can show large differences between individuals that are 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) and individual differences in hydration at the time of testing (for the seminal discussion of these issues, see Lykken & Venables, 1971). Therefore, raw SCL scores are generally not of specific interest, but variation within a particular individual's range is of interest in psychological research. Ben-Shakhar (1985) clarified this further in recommending the use of within-subject standardised scores, as this relies on the more reliable statistic of a mean score.

Following this recommendation, then, we computed standardised scores to reflect departure from the averaged baseline measurements for SCL and HR, and both measures were converted to % 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 5.0uMho that changes to 10.0uMHO represents a 100% increase, whereas an increase from 15.0uMho to 20.0uMho 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 % increase or decrease from the baseline ([score – baseline]*100/baseline). This adjusted score reflects the magnitude of deviation from that individual's unique baseline.

Skin conductance. The initial multivariate analysis indicated a significant increase in SCL over time ($F_{(28,1064)} = 7.85$, $p < .0001$, $\eta^2 = .17$), a significant interaction of condition and sex ($F_{(1,38)} = 4.58$, $p < .03$, $\eta^2 = .11$), and a significant interaction between change in SCL and sex and condition ($F_{(28,1064)} = 2.16$, $p < .0001$, $\eta^2 = .05$), indicating that the trends differed across groups. As expected, girls in the high ST condition showed substantially stronger increases (2.9 times that of the boys and 4.1 times that of the low ST girls; see Figure 2).

Heart rate. The analyses of heart rate yielded no statistically significant effects, although trends were in the expected direction.

Surface skin temperature. There were three subjects whose data on this channel were lost due to a sensor failure. Note that surface skin temperature reacts more slowly to stimuli than either HR or SCL, and has a more restricted reaction range due to biological necessity. Core body temperature, and by association surface skin temperature,

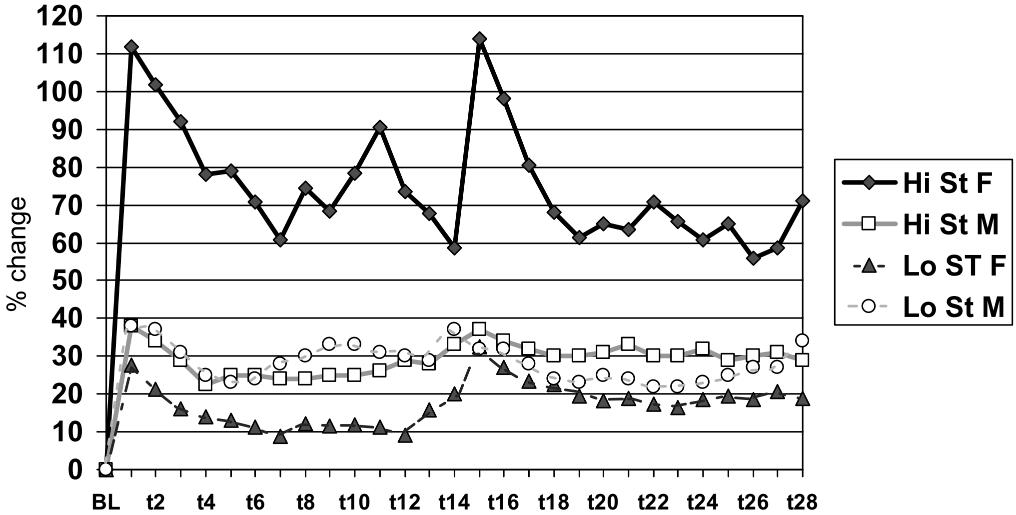

Figure 2. Percentage change from baseline SCL temperature (60-second intervals, reported in °C; monitoring for the first test ended at t14 and monitoring for the second test began at t15)

cannot vary for long outside a relatively narrow given range and still maintain life-supporting processes.

The initial multivariate analysis indicated a highly significant and strong effect for TEMP changes over time as a function of sex and condition only ($F_{(13,546)} = 13.59$, $p < .0001$, $\eta^2 = .24$), with no other significant effects (all $p > .20$). Results controlled for initial baseline TEMP and identification with academics, and are presented in Figure 3. As predicted, there is a clear decrease in TEMP for girls in the high ST group, while the other groups show either no substantial change or increasing TEMP. Given that skin temperature is a relatively slow-changing and constrained variable, it is not surprising that initial observations show little difference across groups.

Blood pressure. Blood pressure difference scores, for both systolic (SBP) and diastolic (DBP) readings, were converted to % change scores between the reading taken at the mid-point of the tests and the baseline. Because there was only one change score per participant, the analysis of these variables was a univariate ANOVA utilising the planned contrast as described above.

The results of these analyses indicated no significant difference across groups for systolic blood pressure, but for diastolic blood pressure there was a highly significant effect. As expected, girls in the high ST condition showed dramatic increases in DBP (8.22%) from baseline, while girls in the low ST condition showed little change (1.63% increase), and boys in the high and low ST conditions showed DBP decreases (-2.30% and –5.89%, respectively; $t_{(36)} = 2.50$, $p < .01$, $d = .83$). These findings are congruent with the temperature readings and the hypothesis regarding blood vessel constriction.

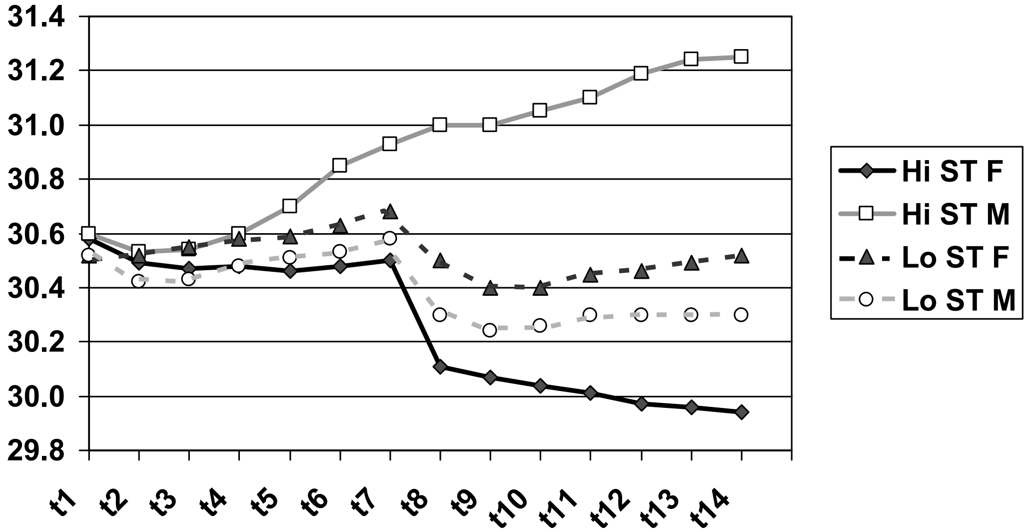

Figure 3. Change in surface skin temperature (60-second intervals, reported in °C; T1–T7 represents the first test and T8–T14 represents the second test; some time elapsed while blood pressure was taken)

Discussion

This study was designed to explore Steele's hypothesis that students subjected to stereotype threat experience relatively inflated levels of anxiety, stress, or arousal. Consistent with expectations, the results of this study show evidence of physiological reactance in girls under high stereotype threat conditions that are consistent with an anxiety or autonomic arousal reaction. Substantial and significant reactance 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 of Blascovich et al. (2001).

As noted above, there are good theoretical and empirical reasons to believe that increased physiological arousal or anxiety could impair cognitive performance, particularly on challenging or unfamiliar tasks. This study fills a gap in the stereotype threat literature by attempting to explicate the mechanism through which stereotype threat manipulations achieve the performance effects noted in many other studies.

This study is not ideal, however, as the participants were university students, and therefore represent a select segment of the student population most likely to be academically successful. However, as many of these participants were psychology majors, many of whom are notoriously maths-averse, it may not be as difficult a generalisation as one might expect. Additionally, if this sample represented only the most successful segment of the student population, it would 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.

Other unanswered questions include the mechanism by which the arousal or anxiety is impairing performance it seems int. Psychological research going back decades suggests that when an individual is experiencing autonomic arousal or anxiety, the result should be decreased memory performance, attention deficits, and processing impairment, particularly on challenging, cognitively-intense tasks. Another direction for future research would be to assess these variables in the context of the stereotype threat paradigm.

Caveats

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. Labelling 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 maths performance). Although it seems intuitive to generalise to other instances (e.g., race, age), that should be done with care, as there is little evidence supporting such a generalisation.

Furthermore, it is reasonable to wonder whether experimenter sex may have had an influence on the reactions of the participants (following research by Inzlicht & Ben-Zeev, 2000, and others). In this study half the subjects experienced a male experimenter, half a female experimenter. Both experimenters did half the subjects in each cell to control for the possibilities, but 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 maths teacher can reduce stereotype threat, or perhaps that a teacher of colour can reduce the threat felt by students of colour? A finding of that nature would be interesting.

Finally, as with much of this type of research, individual responses varied tremendously, hurting statistical inference through increased within-cell variance. Some individuals in the high stereotype threat condition showed virtually no reaction, while others responded dramatically. In an attempt to control for these individual differences, student identification with academics (a necessary ingredient in Steele's model) was measured and covaried in all analyses.

This field and this paradigm need to move to a more individualistic approach to understanding the effects of stereotypes at an individual level. Future research needs to look at what predicts whether an individual will be resilient in the face of negative stereotypes, for example, and how to encourage non-resilient students to be more resilient. To date, 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, as noted above, identification with academics (e.g., Steele, 1997).

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 question high-stakes testing in education carefully.

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 minimise stereotype threat, or maximise student resilience in the face of ubiquitous negative group stereotypes. It has now been over a decade since Steele's *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.

Acknowledgements

I would like to thank Chris Walker, Max Simmons, and Jacklyn Starrett for help with data collection. Some of these data were collected while I was on the faculty of, and with support from, the University of Oklahoma.

Notes

1. Note that although many of these studies claim to manipulate "stereotype threat" there is no direct measure of stereotype threat used in these studies, so this claim is based on inference.
2. In the case of some physiological measurement, there are occasionally missed or dropped measurements due to occurrences such as participant finger movement. These are fairly rare, and in this study were primarily missed heart rate measurements.

References

Alexander, K. L., & Entwhistle, D. R. (1988). Achievement in the first two years of school: Patterns and processes. *Monographs of the Society for Research in Child Development, 53,* 1–157.

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,* 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* (pp. 83–103). 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, Vol. 2: The impact of family background and intelligence on 10th-grade boys.* Ann Arbor, MI: 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.

Ben-Shakhar, G. (1985). Standardization within individuals: Simple method to neutralize individual differences in skin conductance. *Psychophysiology, 14,* 408–413.

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,* 225–229.

Bradley, M. M. (2000). Emotion and motivation. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 602–642). 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 medicational 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,* 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 ed., pp. 224–264). 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 ed., pp. 200–223). 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,* 345–355.

Derakshan, N., & Eysenck, M. W. (1998). Working memory capacity in high trait-anxious and repressor groups. *Cognition and Emotion, 12,* 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: 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,* 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.

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,* 449–455.

Osborne, J. W. (1997). Race and academic disidentification. *Journal of Educational Psychology, 89,* 728–735.

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

Osborne, J. W. (2004). Identification with academics and violence in schools. *Review of General Psychology, 8,* 147–162.

Osborne, J. W., & Rausch, J. L. (2001, April). *Identification with academics and academic outcomes in secondary students.* Paper presented at the American Education Research Association, Seattle, WA.

Quinn, D., & Spencer, S. J. (1996, August). *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: 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,* 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,* 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.

Tobias, S., & Weissbrod, C. (1980). Anxiety and mathematics: An update. *Harvard Educational Review, 50,* 63–80.

Valencia, R. R. (2002). The plight of Chicano students: An overview of schooling conditions and outcomes. In R. R. Valencia (Ed.), *Chicano school failure and success: Past, present, and future* (pp. 3–51). New York: Routledge Falmer.

Voelkl, K. E. (1996). Measuring students identification with school. *Educational and Psychological Measurement, 56,* 760–770.

Voelkl, K. E. (1997). Identification with school. *American Journal of Education, 105,* 294–318.

Whitworth, R. H., & Barrientos, G. A. (1990). Comparison of Hispanic and Anglo graduate record examination scores and academic performance. *Journal of Psychoeducational Assessment, 8,* 128–132.

Wigfield, A., & Eccles, J. S. (1989). Test anxiety in elementary and secondary school students. *Educational Psychologist, 24,* 159–183.

Zuckerman, M., & Lubin, B. (1965). *Manual for the MAACL.* San Diego, CA: Educational and Industrial Testing Service.