

Journal of Black Psychology

<http://jbp.sagepub.com/>

The Effects of Stereotypes on the Achievement Gap: Reexamining the Academic Performance of African American High School Students

J. Thomas Kellow and Brett D. Jones
Journal of Black Psychology 2008 34: 94
DOI: 10.1177/0095798407310537

The online version of this article can be found at:
<http://jbp.sagepub.com/content/34/1/94>

Published by:



<http://www.sagepublications.com>

On behalf of:



Association of Black Psychologists

Additional services and information for *Journal of Black Psychology* can be found at:

Email Alerts: <http://jbp.sagepub.com/cgi/alerts>

Subscriptions: <http://jbp.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://jbp.sagepub.com/content/34/1/94.refs.html>

The Effects of Stereotypes on the Achievement Gap: Reexamining the Academic Performance of African American High School Students

J. Thomas Kellow
Mercer University

Brett D. Jones
Virginia Tech

This study investigated whether African American high school freshman students experience stereotype threat when taking a test that is seen as a predictor of their success on a high-stakes test. The authors conceptually replicated a previous study by Kellow and Jones (2005) using a true experimental design, as opposed to a quasi-experimental design. White students scored statistically significantly higher than African American students when told that their test performance would be predictive of their performance on a statewide, high-stakes standardized test. The following potential mediators of stereotype threat were included in the study and the effects of each are discussed: (a) perceptions of ability and expectancy for success, (b) achievement goal orientation, (c) anxiety, and (d) perceptions of stereotype threat. The results suggest that African American students are at a disadvantage compared to White students when their knowledge and skills are measured using a high-stakes standardized test. Implications for high-stakes testing of African American students are discussed.

Keywords: *stereotype threat; achievement gap; African American academic achievement; stereotype lift*

Stereotypes related to the intellectual performance of African Americans are pervasive in American society (Devine, 1989). Stereotypes about individuals based on group membership persist in part because they often are true for at least some individuals in the group. For instance, it is true that

African American students, in general, score lower on standardized academic achievement tests than White students (Steele, 1997). Television news reports and newspaper articles reinforce these stereotypes with discussions about the achievement gap between African American and White students (see Tobin, Ave, & Batts, 2004; Tobin & Batts, 2004). These media reports likely perpetuate negative stereotypes about African American students' abilities to succeed on standardized tests and in school.

Stereotypes related to testing are especially important to consider in this age of accountability given the unprecedented focus on standardized tests as a means to gauge student progress. In fact, much of what educators and students do in public schools today is related to preparation for or in response to standardized testing (Jones, Jones, & Hargrove, 2003). Much of the focus on testing is the result of the No Child Left Behind Act of 2001 (NCLB, 2002), which has forced states to implement statewide accountability systems that require annual testing in Grades 3 through 8. Furthermore, NCLB requires that these test results be subdivided by poverty, race, ethnicity, disability, and limited English proficiency to ensure that none of these groups are left behind (U.S. Department of Education, 2001). This subdivision of test scores places a focus on the differences between groups of students, including the large gap in achievement scores between African American and White students. The widely held negative stereotypes related to the intellectual performance of African Americans and the increased reliance on standardized tests as a means to measure student progress create a set of conditions that would appear conducive to African Americans' experiencing *stereotype threat* during testing.

STEREOTYPE THREAT

Stereotype threat is the threat that members of a stigmatized group experience when they believe that they may, by virtue of their performance in a domain of relevance, confirm a negative stereotype about themselves and members of their group. This phenomenon was first studied by Steele and Aronson (1995), who proposed that differences in academic performance between minority and nonminority students, as measured by standardized achievement tests such as the SAT, could partially be explained by anxiety and evaluation apprehension produced by knowledge of negative stereotypes

AUTHORS' NOTE: Correspondence concerning this article should be directed to J. Thomas Kellow, PhD, Mercer University, 1408 Westchester Ridge NE, Atlanta, GA 30329; e-mail: tkellow@pdq.net or kellow_jt@mercer.edu.

related to group membership. Consistent with this hypothesis, they found that when a task was presented to African American college students as indicative of verbal academic ability, they performed far worse than a matched group of students who were told the identical task measured psychological processes involved in verbal problem solving. Similar results have been obtained by presenting a measure as a “traditional” test of achievement or intellect or as a “culture-free” or “nonbiased” test (Mayer & Hanges, 2003). In fact, researchers have found that simply indicating one’s race prior to taking an achievement test was sufficient to activate stereotype threat (Steele, 1997).

We were interested in whether the phenomenon of stereotype threat could partially explain the achievement gap between White and African American high school students. Although the effects of stereotype threat on test performance have been well documented at the college level (see Steele, Spencer, & Aronson, 2002, for a review), few studies have specifically measured the effects of stereotype threat on the test scores of K-12 students (e.g., Keller, 2002). Given the unprecedented use of high-stakes testing in K-12 schools to determine grade promotion and graduation, this population deserves attention.

This study examines the extent to which African American high school freshman students experience stereotype threat when taking a test that is seen as a predictor of their success on a high-stakes test. We chose this population for several reasons. First, many high school students across the country must pass high-stakes standardized tests in various content domains to obtain a diploma. Students transitioning from middle to high school are often explicitly or implicitly reminded of this fact by parents, teachers, administrators, and their peers. Yet little is known about whether stereotype threat is a factor in the test scores of high school students. Second, these standardized tests are typically given at the 10th- and sometimes the 11th-grade level. If stereotype threat is indeed a phenomenon that exists with these students, the 9th-grade would seem to be a minimum starting point for interventions designed to reduce the threat (although ideally efforts would begin well before this point).

DEPENDENT VARIABLES OF INTEREST

Because researchers (e.g., Smith, 2004) have proposed a number of potential mediators of stereotype threat, we included the following ones in this study: (a) perceptions of ability and expectancy for success, (b) achievement goal orientation, (c) anxiety, and (d) perceptions of stereotype threat. A description of each of these variables is provided in this section.

PERCEPTIONS OF ABILITY AND EXPECTANCY FOR SUCCESS

Eccles and her colleagues (Eccles, Adler, & Meece, 1984; Eccles et al., 1983; Eccles & Wigfield, 1995; Wigfield & Eccles, 1992) have developed an expectancy-value model of motivation that predicts that student performance is directly affected by both expectancies and values. They have tested their model empirically with elementary through secondary school students in mathematics and English and have found that students' self-perceptions of ability and expectancies for success relate strongly to their achievement (Eccles, 1984a, 1984b; Eccles et al., 1983; Meece, Wigfield, & Eccles, 1990) and their use of more effective cognitive and metacognitive strategies (Pintrich, 1989, 1999; Shell, Murphy, & Bruning, 1989). Students' self-perceptions of ability and expectancies for success vary by academic domain (e.g., mathematics, English, science, social studies) and are based on the confidence they have in their abilities and on their estimation of the difficulty of the domain (Pintrich & Schunk, 2002). Furthermore, expectancy beliefs can vary within a domain depending upon the task and context (Schunk & Pajares, 2005). To capture the possible difference in self-perceptions between the domain and a particular task within the domain, we measured students' self-perceptions of ability and expectancies for success in the domain of mathematics and their proximal expectancy for success on a specific mathematical task (a spatial ability task).

Expectancy for success is an important variable to consider in studies related to stereotype threat because it has been hypothesized to be a possible mediator of stereotype threat (Aronson & Steele, 2005). This hypothesis is based on the finding that task expectancies can be influenced by one's perceptions about the stereotypical characteristics of the task (Stangor, Carr, & Kiang, 1998). Therefore, we would expect that when African American students are placed in a testing situation designed to invoke negative stereotypes, their expectancies for success would be lowered.

ACHIEVEMENT GOAL ORIENTATION

Achievement goal theory examines how and why students approach and engage in achievement tasks (Pintrich & Schunk, 2002). In broad terms, goal orientations can be divided into mastery and performance (Dweck, 1986; Nicholls, 1984). Mastery goals are focused on learning and on the development of competence through task mastery. Performance goals are focused on the demonstration of competence relative to others. Students who focus on learning and mastering a task are described as having a mastery-approach goal, whereas students who focus on avoiding not-learning have a mastery-avoidance goal (Elliot & McGregor, 2001). Similarly, students

who focus on being the best and smartest in comparison to others have a performance-approach goal, and students who focus on avoiding looking incompetent have a performance-avoidance goal (Elliot & Church, 1997). In general, researchers have found students with mastery goal orientations to be more engaged in learning than students with performance-avoidance goal orientations (Wolters, 2004).

Recently, researchers have hypothesized that a testing environment that induces stereotype threat may orient students to a particular achievement goal. Ryan and Ryan (2005) have hypothesized that situational cues that induce stereotype threat serve to focus an individual towards a performance-avoidance goal. Similarly, Smith (2004) has proposed a stereotyped task engagement process (STEP) model that predicts that stereotype threat triggers a performance-avoidance goal. These researchers make these assertions based on the idea that students experiencing stereotype threat want to prove that the stereotype is not relevant to their academic competence. Smith further hypothesized that a nullified stereotype engenders a performance-approach goal. Although no empirical research exists at this time to confirm the relationship between stereotype threat and achievement goals, the theoretical base behind these hypotheses appears strong. In this study we examine this hypothesis.

ANXIETY

A number of studies (see Smith, 2004; Steele & Aronson, 1995) have examined the possibility that stereotype produces feelings of anxiety. These researchers hypothesize that fear of confirming a negative stereotype related to one's group membership (e.g., African American) elicits an anxiety response that in turn produces cognitive interference that undermines test performance. Anxiety has been operationalized in a variety of ways, including (a) self-report inventories, (b) physiological measures such as blood pressure, and (c) word fragment tests designed to elicit negative adjectives related to the stereotyped group (Mayer & Hanges, 2003; Smith, 2004). Results from studies that have examined the role of anxiety in mediating stereotype threat are mixed, with some studies demonstrating an experimental effect on anxiety and others producing no results.

In this study, we chose to assess the role of anxiety because it may have a more profound effect on adolescents than adults. This assertion is based on an adolescent developmental phenomenon known as the "imaginary audience." Numerous studies (see Vartanian, 2000, for a review) have documented the existence of the imaginary audience, wherein adolescents tend to believe they are constantly being evaluated by others with respect to their personal characteristics (e.g., ethnicity and intelligence) and behavior. This preoccupation with the way in which others evaluate them often enhances

anxiety when the person is called upon to perform a task, even if that performance is private or not directly observable. This would seem to make performance on a test that is presumably diagnostic of one's ability particularly anxiety provoking for this population. Therefore, we would expect African American students in a stereotype threat condition to be more anxious than those in a less threatening condition.

PERCEPTIONS OF STEREOTYPE THREAT

Pinel (1999, 2002) has suggested that *stigma consciousness*, "the extent to which a person expects to be stereotyped by others" (Ployhart, Ziegert, & McFarland, 2003, p. 238), influences reactions that a stereotyped person may have in a number of situations (e.g., interpersonal; test-taking). It is important to note that the extent to which a person believes he or she is stereotyped by others in a given domain does not imply that the person *believes* the stereotype to be true; rather, it simply indicates the individual's awareness of the stereotype. However, heightened perceptions of the existence of a stereotype related to one's group membership may exacerbate the potential for *evaluation apprehension* (Steele, 1997) and make one more vulnerable to the fear of confirming the expectations of others regarding a group stereotype (Steele et al., 2002). Ployhart et al. (2003) have suggested two distinct types of perceptions regarding the prevalence of group stereotypes: (a) general and (b) specific. General stereotype perceptions involve rather global assessments of the extent to which a person is stereotyped across a variety of domains. In contrast, specific stereotype perceptions are contextually dependent and refer to perceptions related to a specific task or performance domain (e.g., achievement test performance).

Although relatively unexplored, it appears that this variable is most prominent in explaining stereotype threat effects when directed towards a specific task (Mayer & Hanges, 2003). We would expect that inducing stereotype threat in a testing situation would lead African American students to report higher levels on a stereotype threat-specific scale, especially when compared to a group of African American students who are told that the test is unbiased with respect to gender and ethnicity.

RATIONALE FOR THE STUDY

There are several important limitations in the current body of research on stereotype threat. First, of the more than 40 empirical studies conducted with minority participants, all but 1 have used college students and other

adult samples. Only 2 studies of which we are aware have been conducted with high school students. One (Stricker, 1998) employed a very select sample (i.e., those taking the Advanced Placement Calculus Examination). The second study (Kellow & Jones, 2005) utilized a more representative population of high school freshman and focused on the overall stereotype threat effect, as well as anxiety and perceptions of ability and expectancy for success. Although the findings from the latter study were promising, their use of intact groups and a subsequent analysis of covariance (ANCOVA) procedure to control for initial group differences in ability based on previous test scores is potentially problematic. Wicherts (2005) has pointed out several limitations of using quasi-experimental designs and concomitant ANCOVA adjustments. These problems involve violations of ANCOVA assumptions, most notably (a) homogeneity of regression and (b) covariation of the treatment and the pretest variable. It is beyond the scope of this article to explore these concerns in detail, but the result of these ANCOVA violations, if present, is a distortion of adjusted means for cell groups and an increase in Type I error rates: "In plain English, saying things that aren't so" (Cohen, 1990, 1304).

We expand on the Kellow and Jones (2005) study in two important ways: (a) using a randomized experimental design and (b) including measures of achievement goal orientation, perceptions of stereotype threat-specific, and a more proximal measure of expectancy for success. Importantly, we used the same stimulus materials as Kellow and Jones, including the same directions and performance outcome measure.

RESEARCH HYPOTHESES

Each of the following hypotheses includes a statement about an "evaluative condition," a "nonevaluative condition," or both. The "evaluative condition" refers to the group of students who were told that their test scores on the test that they were about to take would predict how well they would score on a standardized high-stakes test. The "nonevaluative condition" refers to the students who were told that the test they were about to take was unbiased.

Hypothesis 1: African American participants in the evaluative condition will score lower on a measure of visual spatial ability than White participants in the same condition.

Hypothesis 2: African American participants in the evaluative condition will report lower self-perceptions of ability and expectancies for success than White participants in the same condition.

Hypothesis 3: African American participants in the evaluative condition will report lower proximal expectancies for success than White participants in the same condition.

Hypothesis 4: African American participants in the evaluative condition will report higher levels of performance avoidance than White participants in the same condition.

Hypothesis 5: African American participants in the nonevaluative condition will report higher levels of performance approach than White participants in the same condition.

Hypothesis 6: African American participants in the evaluative condition will report higher levels of state anxiety than White participants in the same condition.

Hypothesis 7: African American participants in the evaluative condition will report higher levels of stereotype threat-specific than African American participants in the nonevaluative condition.

METHOD

An unbalanced 2×2 ANOVA using no covariates was employed. Prior to the substantive analyses, we tested all required ANOVA assumptions and none were violated. The independent variables were (a) student ethnicity (African American or White) and (b) condition (evaluative or nonevaluative). Dependent variables were the (a) Applied Personnel Research Spatial Ability Test (APR), (b) proximal expectancy for success item, (c) achievement goal orientation scale, (d) Self- and Task-Perception Questionnaire (STPQ), (e) State Trait Anxiety Inventory (STAI), and (f) the stereotype threat-specific scale. All of these variables are described below.

PARTICIPANTS

Participants were recruited from freshman introductory algebra courses at a large urban high school in Florida. Informed consent forms for parent or guardian approval were provided to 641 students. The study was presented as an investigation of mathematical reasoning. For their participation, students received a gift card for a free meal at a local restaurant. One hundred eighteen (18.4%) of the forms were signed and returned.

We obtained participants' ethnicity and gender classifications from the school's database. The ethnic composition for students returning forms was White 53%, African American 42%, Hispanic 3%, and Asian 2%. Sixty-six percent of the students were female. Based on the demographics of all ninth-grade students at this school, African Americans and females were

slightly, but not statistically significantly, overrepresented in the sample. Data on Free/Reduced Lunch (a proxy for socioeconomic status) were not obtained given the questionable validity of such measures for high school students (see Willson & Kellow, 2002). While desirable, other proxy measures, such as parental income or occupation, simply could not be obtained without extraordinary measures. Students were stratified according to ethnicity and then randomly assigned to either the experimental (evaluative) or comparison (nonevaluative) condition. On the day of testing, 11 students (8 African American and 3 White) were absent and 1 student failed to complete the dependent measure. Because the focus of the study was on African American and White students and because so few Hispanic and Asian students participated (combined $n = 5$), data for the Hispanic and Asian participants were excluded from the final sample. As a result, the total number of student participants was 101. All participating students were average to excellent in mathematics performance based on the normal curve equivalent (NCE) scores obtained by each group on eighth-grade standardized tests (White $M = 71.6$; African American $M = 50.1$).

MEASURES

APR. The APR was developed by Wiesen (1996) as a personnel selection test for skilled clerical applicants. It is a timed test (we allowed 5 minutes) consisting of 50 items that relate to a set of blocks that are stacked in various configurations (see Figure 1 for an example item). Respondents must indicate how many other blocks are touched by a specific block. For instance, for the example in Figure 1, respondents indicate the number of blocks touching the block labeled A (the answer is two—one above and one below block A). This type of problem is consistent with the types of knowledge and skills that are expected of ninth-grade students. In fact, the National Council of Teachers of Mathematics (NCTM; 2000) geometry standards specifically state, "In grades 9-12 all students should visualize three-dimensional objects from different perspectives and analyze their cross sections" (p. 308).

Data from the eighth-grade administration of the mathematics Florida Comprehensive Assessment Test (FCAT) were available for 85 of the students in the present study. The correlation between these scores and the APR was .55. Kellow and Jones (2005) performed a modified split-half procedure appropriate for speeded tests (Anastasi, 1988) on a subsample of 24 students and obtained a Spearman-Brown adjusted score reliability of .92. The APR served as the primary dependent variable in the study.

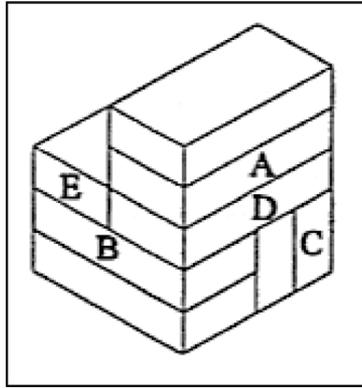


Figure 1: Example Applied Personnel Research Spatial Ability Test (APR) Item

Proximal expectancy for success. Prior to attempting the APR, we measured students' proximal expectancy for success using a single item: "The test you are about to take has 10 block problems similar to the one above. Please circle the one number below that best describes how well you expect to do on this test." Students responded on a 7-point Likert-format scale anchored at 1 (*not well at all*) and 7 (*very well*). The item was modeled after similar items developed by Eccles and Wigfield (1995) that measure expectancies for success at the domain level (e.g., "How well do you think you will do in your math course this year?").

Achievement goal orientation. Items adapted from Elliot and McGregor (2001) were used to assess students' goal orientation during the test. One item was adapted to measure students' performance-avoidance goal orientation, and another was adapted to measure students' performance-approach goal orientation. The performance-avoidance goal item read, "I just wanted to avoid doing poorly on the test I just took"; and the performance-approach goal item read, "It was important for me to do well compared to other students on the test I just took." Students responded on a Likert-format scale anchored at 1 (*not at all true of me*) to 7 (*very true of me*). These items were selected because they could be adapted to discern the goal orientation of students during the test, as opposed to measuring their more general goal

orientation in a class (e.g., “My goal in this class is to get a better grade than most of the other students”).

STPQ. This measure has been used with adolescents to assess a variety of constructs related to their beliefs, values, and attitudes regarding an academic domain, such as mathematics (Eccles & Wigfield, 1995). We used the five-item scale that pertained to ability perceptions and expectancies for success. Each item was scaled using a 7-point Likert-format scale. The reported score represents an average of the five items. A sample item is, “Compared to other students, how well do you expect to do in math this year?” Students responded on a 1 (*much worse*) to 7 (*much better*) scale. Previous researchers (e.g., Eccles et al., 1984; Eccles & Wigfield, 1995) have empirically evaluated the psychometric properties of the scale and report strong factorial validity and reliability of scores ($\alpha = .92$). The correlation between the total score and students’ eighth-grade FCAT mathematics scores in the present study was .25. We conducted an internal consistency analysis (Cronbach’s alpha) on the present data and found a reliability estimate of .83.

STAI. A short-form of the STAI (Spielberger, Gorsuch, & Lushene, 1970) consisting of eight items with a 4-point Likert-format scale anchored at 1 (*not at all*) and 4 (*very much*) was used to assess state anxiety of students while taking the APR. Items were summed to produce a total score. A sample item is, “While taking the test I was tense.” The STAI has a strong psychometric reputation built on numerous empirical studies (see Anastasi, 1988, for a review). In the present study, the STAI and APR scores were modestly negatively correlated ($r = -.10$). In addition, the scale yielded a respectable internal consistency estimate ($\alpha = .84$).

Stereotype threat-specific scale. We measured the extent to which students expected to be stereotyped on the APR based on their ethnicity using a 5-item Likert-format scale anchored at 1 (*strongly disagree*) and 5 (*strongly agree*) and reported the average for the five items (Ployhart et al., 2003). A sample item is, “Others have negative thoughts about how people of my race perform on this type of test.” Internal consistency for the five items was .78.

PROCEDURE

Students were randomly assigned to either an evaluative or nonevaluative condition. Students were tested in groups during the period that they

ordinarily would have their mathematics class. The groups ranged in size from 10 to 22. There were a total of 48 students in the evaluative condition (23 White and 25 African American) and 53 students in the nonevaluative condition (31 White and 22 African American). After arriving at one of two classrooms reserved for the experimenters (who were both White males), participants were provided with a pencil and a booklet containing the student assent form, the APR, and the various instruments described previously. After reading and signing the assent form, students received one of two sets of instructions described below.

Evaluative condition instructions. "You will be taking a test that consists of challenging questions about mathematical reasoning. We are very interested in this test because students who score highly on this test tend to score highly on the 10th grade mathematics FCAT and students who do poorly on this test tend to do poorly on the 10th grade mathematics FCAT. You will have 5 minutes to complete this mathematical reasoning test. It is important that you do your very best, because we will use your scores to give you information about your strengths and weaknesses that will help you pass the 10th grade mathematics FCAT."

Nonevaluative condition instructions. "You will be taking a test that consists of challenging questions about mathematical reasoning. We are very interested in this test because research has shown that boys and girls score the same on it. Research has also shown that students of different ethnicities, such as White, Black, or Hispanic students, score the same. We call this test a gender and culture fair test because it is unbiased. You will have 5 minutes to complete this mathematical reasoning test. It is important that you do your very best, because we will use your scores to give you information about your strengths and weaknesses that will help you pass the 10th grade mathematics FCAT."

Following the instructions, students were guided through an example item to ensure that they understood the proper procedure for interpreting and responding to the items. Next, students completed the item measuring their proximal expectancy for success. Students were then told, "When you are told to begin, go to the next page and complete as many of the problems as you can. Please do not skip any of the problems. Give your best answer. You only have 5 minutes, so you must work quickly." Students were then told to begin and they worked on the APR. After 5 minutes, students were told to stop (whether they had attempted all of the items or not) and begin to complete the questions on the following pages, which contained the

achievement goal orientation items, the STAI, the STPQ, and the stereotype threat-specific scale. When students finished, they handed in their answers and returned to their mathematics class.

Soon after the experiment, the researchers revisited the school and debriefed students as to the true purpose of the study as well as answered any questions the students had pertaining to the experiment. In addition, they were provided with a debriefing sheet to share with their parent or guardian.

REPORTING AND ANALYSIS

For all analyses, we set p critical at .05. We report exact p values for all results, supplemented by partial eta-squared (η^2_{partial}) as an uncorrected effect size measure. In addition, 95% confidence intervals based on non-central F distributions were calculated for the effect size estimates using an SPSS routine developed by Smithson (2001). Statistical power is presented as $(1 - \beta)$, where β is the probability of committing a Type II error for a given result. Overall model fit was assessed using unadjusted R^2 . All analyses were conducted with SPSS 12.1 using the General Linear Model procedure using adjusted Type III sums of squares to compensate for the unbalanced nature of the design.

RESULTS

On the APR, the primary dependent variable in this study, participants attempted an average of 27.8 items ($SD = 7.1$) out of the 50 possible items. However, they only successfully completed an average of 16.7 items ($SD = 7.0$), for an average of 60.1% correct on the attempted items. This finding suggests that the participants found the test challenging, which is important to consider in this study because stereotype threat is only triggered during challenging tasks (Spencer, Steele, & Quinn, 1999).

HYPOTHESIS 1

Hypothesis 1 was that African American participants in the evaluative condition would score lower on the APR than White participants in the same condition. The results of the experimental manipulation on this variable are presented in Table 1, and the interaction effect is depicted in Figure 2.

TABLE 1
Student Performance on the Applied Personnel Research Spatial Ability Test (APR) by Ethnicity and Condition

<i>Effect</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	<i>Partial η^2</i>	<i>95% Confidence Interval</i>	<i>Power (1 - β)</i>
Ethnicity	1	1,205.95	33.89	.00	.26	.12 to .38	.99
Condition	1	149.41	4.21	.04	.04	.01 to .14	.54
Ethnicity \times Condition	1	157.71	4.43	.04	.04	.01 to .14	.54
Error	97	3,451.22					
Model $R^2 = .29$							

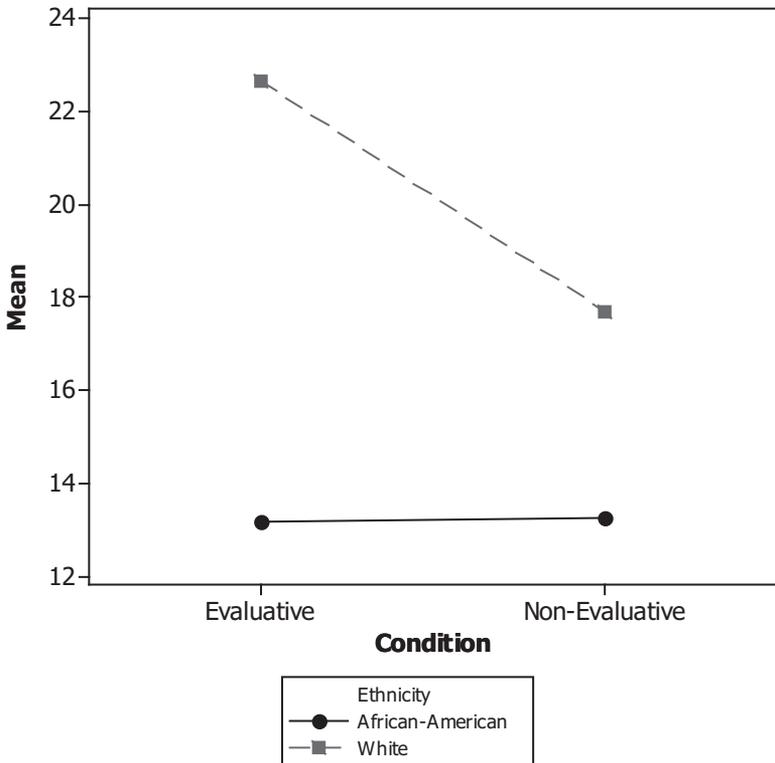


Figure 2: Interaction of Ethnicity and Condition on Student Applied Personnel Research Spatial Ability Test (APR) Performance

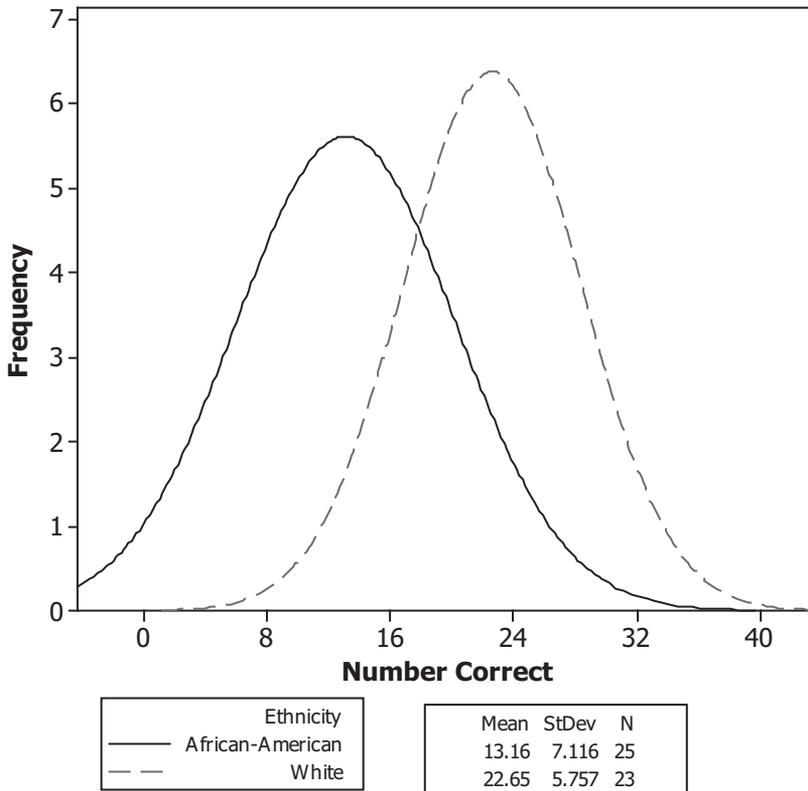


Figure 3: Mean Number of Applied Personnel Research Spatial Ability Test (APR) Items Correct in the Evaluative Group by Ethnicity

While the interaction effect was statistically significant (as predicted), an interesting pattern of performance emerged. African American students in the evaluative condition ($M = 13.16$) performed almost identically to African American students in the nonevaluative condition ($M = 13.23$). With respect to White students, however, there was a sharp difference between conditions, with students in the evaluative condition ($M = 22.65$) outperforming students in the nonevaluative condition ($M = 17.68$). This difference is displayed in Figure 3. It should be noted that the standardized difference effect size d (Cohen, 1988) in the evaluative condition between African American and White students ($d = 1.4$) is similar to that found by Kellow and Jones (2005) ($d = 1.3$).

HYPOTHESES 2 AND 3

With respect to the second hypothesis, we predicted that African American students in the evaluative condition would report lower self-perceptions of ability and expectations for success (as measured by the STPQ) than White students in the same condition. In the evaluative condition, the mean of African American participants on the STPQ ($M = 5.33$) was slightly higher than that of White participants ($M = 4.98$). Similarly, African American students in the nonevaluative conditions had a slightly higher mean ($M = 5.11$) than White students ($M = 5.03$). None of these differences, however, were statistically significant at the prescribed alpha level.

Our third hypothesis was similar to our second, except that instead of measuring students' perceptions of ability and expectancies for success in the domain of mathematics, we measured their proximal expectancy for success specifically on the APR that they were about to complete. The results are presented in Table 2, and the interaction effect is presented in Figure 4. African American and White students in the nonevaluative condition had nearly identical proximal expectancy for success on the APR (African American $M = 4.77$; White $M = 4.81$). However, in the evaluative condition, African American participants had a statistically significantly lower mean proximal expectancy for success ($M = 4.04$) than White participants ($M = 5.09$).

HYPOTHESES 4 AND 5

Regarding achievement goal orientation, we predicted that African American students in the evaluative condition would have higher performance-avoidance goals and lower performance-approach goals than White students. However, there were no statistically significant effects associated with either dependent variable. With respect to performance avoidance, African American students in the evaluative condition ($M = 4.76$) were very similar to White students ($M = 4.83$) in their responses. Similarly, African American students in the nonevaluative condition ($M = 4.86$) were very close to White students ($M = 5.13$) in terms of mean differences.

The data related to performance approach mirrored those of the performance-avoidance variable. In the evaluative condition, African Americans ($M = 4.84$) were very similar to White students ($M = 4.69$) in their responses. In the nonevaluative condition, the average for African American students ($M = 4.46$) was close to that of White students ($M = 4.65$).

HYPOTHESIS 6

We hypothesized that African American students in the evaluative condition would experience more state anxiety than White students in the same

TABLE 2
Expectancy for Success by Ethnicity and Condition

<i>Effect</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	<i>Partial η²</i>	<i>95% Confidence Interval</i>	<i>Power (1 - β)</i>
Ethnicity	1	7.25	4.64	.03	.05	.01 to .14	.64
Condition	1	1.27	0.81	.37	.01	.00 to .07	.17
Ethnicity × Condition	1	6.37	4.01	.04	.04	.01 to .14	.54
Error	97	1.56					
Model <i>R</i> ² = .089							

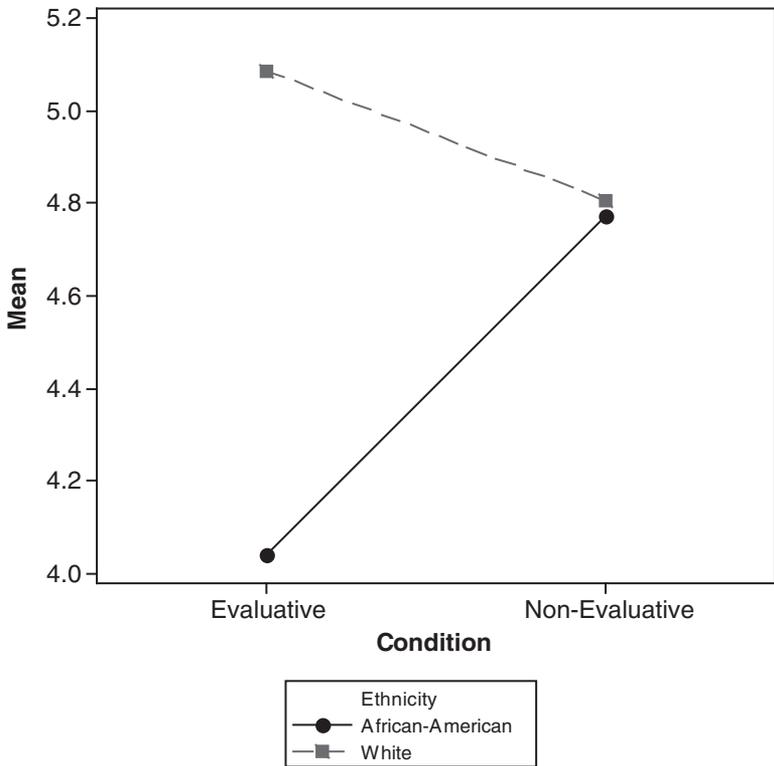


Figure 4: Interaction of Ethnicity and Condition on Proximal Expectancy for Success

TABLE 3
Stereotype Threat-Specific by Ethnicity and Condition

<i>Effect</i>	df	MS	F	p	<i>Partial</i> η^2	<i>95%</i> <i>Confidence</i> <i>Interval</i>	<i>Power</i> <i>(1 - β)</i>
Ethnicity	1	35.01	38.45	.00	.28	.14 to .41	.99
Condition	1	0.99	1.09	.29	.01	.00 to .08	.17
Ethnicity \times Condition	1	0.63	0.69	.41	.00	.00 to .07	.00
Error	97	0.91					
Model $R^2 = .29$							

condition. We found statistically insignificant differences between condition and ethnicity. In the evaluative condition, African American students reported slightly more state anxiety ($M = 15.52$) than White students ($M = 14.13$); however, given the range of possible scores (8 to 32) this difference is clearly negligible. Similarly, in the nonevaluative condition, African American students also scored slightly higher ($M = 15.14$) than White students ($M = 14.81$).

HYPOTHESIS 7

The measure of stereotype threat-specific did not yield the expected interaction effect between ethnicity and group; however, a statistically significant and substantive main effect was noted. African American students' scores in both the evaluative ($M = 2.57$) and nonevaluative ($M = 2.93$) conditions were higher than those of White students in the evaluative ($M = 1.54$) and nonevaluative ($M = 1.58$) conditions (see Table 3).

This finding indicates that African Americans students in both conditions were more likely than White students to believe that they were being stereotyped by others.

DISCUSSION

Perhaps the most interesting finding in this study was the lack of degradation in performance of African American students in the evaluative condition and the profound performance increase of White students in the same condition. This occurred despite the fact that African American students had significantly lower proximal expectancies for success in the evaluative

condition than in the nonevaluative condition. We sought answers to this inconsistency by carefully comparing our sample characteristics to that of Kellow and Jones (2005) with no success. We reviewed the fidelity of the experimental manipulation, explanation of the APR, and testing conditions, yet were left with no clues. Perhaps there were other variables that were not measured in this study that account for the discrepancy, such as student characteristics or the school environment. We offer one tentative explanation for the lack of a significant difference between African American students in the evaluative and nonevaluative conditions on the APR task. A potentially critical difference between the current study and that of Kellow and Jones was the timing of the study. The present study was conducted *after* these ninth-grade students had taken the high-stakes FCAT examination, whereas the Kellow and Jones study was conducted roughly one month *before* students took the FCAT. Is it possible that, for many of the African American students in the evaluative condition, there was an absence of pressure to perform? We offer this only as speculation, but intuitively it seems to make sense that, absent the impending prospect of a high-stakes assessment, the pressure of confirming a negative stereotype about one's group may be diminished.

It may also be possible that the in-group (i.e., White students) may have experienced *stereotype lift*. Stereotype lift occurs when performance is enhanced due to an awareness that an out-group is negatively stereotyped (Walton & Cohen, 2003). In other words, performance is enhanced through a downward social comparison with the out-group. In the present study, White students possibly knew that African American students as a group do not score as highly on the FCAT as White students. White students may have experienced stereotype lift even though there was no explicit reference to an out-group. These results are consistent with the findings of Walton and Cohen (2003), who found that Whites link intellectual tests to negative stereotypes automatically. Furthermore, Walton and Cohen believe that stereotype lift is driven by negative stereotypes about the out-group and not by positive stereotypes about the in-group (the White students in the present study). Clearly, there is much to learn about the impact of high-stakes testing, the testing environment, and the general school climate as it relates to high-stakes testing on the performance of both minority and nonminority students.

The findings related to perceptions of ability and expectancies for success are consistent with Hypothesis 3, but not Hypothesis 2. That is, African American students reported lower proximal expectancies for success than White students (in the same condition) on the task-specific item (related to their spatial ability), but did not report lower self-perceptions of ability and expectancies for success in the domain of mathematics. This finding

suggests that the proximal measure of expectancy for success at the task level is a more sensitive measure than a domain-level measure of abilities and expectancies. This finding appears to be consistent with research that found that expectations are influenced by the stereotypical characteristics of the task or the situation in which the task is to be performed (Stangor et al., 1998). Moreover, these results lend support to Steele's (1997) concept of stereotype threat in which the threat of negative stereotypes is strictly situational and "is cued by the mere recognition that a negative group stereotype could apply to oneself in a given situation" (p. 617). In the present study, we wanted to cue the stereotype that African Americans perform worse on mathematics tests than White students by telling students in the evaluative condition that the results of the APR were predictive of how well they would do on the mathematics FCAT. These findings suggest that relating the test to the mathematics FCAT was sufficient to cue such stereotypes. We did not have to describe the FCAT or provide statistics about how African Americans typically perform on it. This indicates that freshman African American high school students are aware of the negative stereotypes related to this statewide standardized test and that these stereotypes can affect their expectations for success on a particular test.

The results of this study do not support Hypothesis 4: that the evaluative condition triggers a performance-avoidance goal for African American students. This result is important because it is the first empirical test of which we are aware to test the models proposed by Smith (2004) and Ryan and Ryan (2005) that link stereotype threat to the achievement goal research. The theory and rationale behind these models is that the situational cues that incur stereotype threat trigger a performance-avoidance goal because performance-avoidance goals "are conceptually similar to the state in which individuals feel threatened to prove that a competence-stereotype is not valid" (Smith, 2004, p. 195).

As to why the evaluative condition did not trigger a performance-avoidance goal for African American students, we offer a few explanations. First, we were only able to include one item for each goal orientation from the three used by Elliot and McGregor (2001) on the questionnaire related to students' performance goals because the other items did not make sense when rewritten to refer to the APR that they had just completed. Incorporating more items to assess these constructs would certainly be desirable in future studies. Further research should be conducted on whether this type of change in item wording results in valid scores to measure students' performance goals on a test. Furthermore, it is possible that stereotype threat does not trigger a performance-avoidance goal or that there are confounding variables that we did not consider in this study.

In considering the performance-approach goals in Hypothesis 5, Smith's (2004) STEP model hypothesizes that nullifying the stereotype might engender a performance-approach goal. In the present study, we nullified the stereotype in the nonevaluative condition by telling students that the APR is unbiased and that students of different ethnicities score the same on it. African American students in the nonevaluative condition, however, did not report higher values on the performance-approach item than the White students. This result does not support our hypothesis or that proposed by Smith. The possible explanations we discussed related to the performance-avoidance goal in the previous paragraph might also account for the lack of significant findings related to students' performance-approach goal on the test.

For Hypothesis 6, the lack of a statistically significant interaction effect between experimental condition, ethnicity, and the anxiety variable was not entirely unexpected given the mixed findings of previous studies. Interestingly, two studies that provided significant evidence of the role of anxiety used physiological measures of anxiety as opposed to self-report measures. Blascovich, Spencer, Quinn, and Steele (2001) noted increases in the blood pressure of African American college students when placed in a high stereotype-threat condition, while Osborne (2007) found heightened physiological arousal among college females confronted with a challenging mathematics test on which gender differences favoring males on the test were expressly stated. This last study is notable in that multiple measures of arousal (galvanic skin conductance, heart rate, surface skin temperature, and blood pressure) were employed, all of which have been supported as strong physiological correlates of anxiety (Bradley, 2000). In light of these results, there are two intriguing possibilities for the tendency of studies using self-report measures to show nonsignificant results. First, it may be that the anxiety experienced by participants in evaluative conditions may be sufficiently transient that post hoc measures of reported anxiety are insensitive to group differences. Alternatively, it may be that responses on self-report measures are contaminated by a self-presentation bias exhibited by participants. Since anxiety is generally perceived as an undesirable state, persons in anxiety-producing situations may underestimate the actual level of anxiety they experienced during the test.

We believe that future research should focus more closely on the measurement of anxiety. Indeed, it is hard to imagine that the disruptive effects of stereotype threat could exist without an increase in subjective and objective stress mechanisms. It would be desirable to measure anxiety directly through physiological instruments while participants are taking an examination. Unfortunately, there are two significant obstacles to this approach.

First, the use of (multiple) measures of physiological responses to anxiety-producing situations is economically prohibitive. Second, measuring these responses in the naturally occurring environment of a high-stakes test would seem to threaten the social validity (Trochim, 2004) of experimental findings, since it is hardly realistic that students would routinely be harnessed to various instruments that measure autonomic arousal. Nonetheless, more accurate measures of students' level of anxiety during testing are critical to understanding the affect of anxiety in stereotype threatening situations.

Finally, African American participants in the evaluative condition did not report higher levels on the stereotype threat-specific measure than African American participants in the nonevaluative condition as was predicted in Hypothesis 7. However, the fact that African American students' scores in both the evaluative and nonevaluative conditions were statistically significantly higher than those of the White students in their respective groups suggests that, overall, African Americans expected to be stereotyped based on their ethnicity more than the White students.

The Kellow and Jones (2005) study did not directly measure perceptions of stereotypes related to minorities and academic achievement. They did, however, use a single item asking students to indicate their level of agreement with the statement, "I felt that this test was biased against minority students." The item was intended to measure the extent to which the experimental manipulation was effective in (a) alleviating perceptions of bias in the nonevaluative condition and (b) supporting perceptions of bias in the evaluative condition. In retrospect, an examination of several of the stereotype threat-specific items we adopted from Ployhart et al. (2003) could plausibly be construed as different facets of an overall construct of test bias (e.g., "Tests, like the one I just took, have been used to discriminate against people of my race"). This is supported by a pattern of results that mirror those of Kellow and Jones, although the main effect for ethnicity was considerably stronger ($\eta^2_{\text{partial}} = .28$) than the effect noted in the previous Kellow and Jones study ($\eta^2_{\text{partial}} = .11$).

CONCLUSION AND IMPLICATIONS

Although the phenomenon of stereotype threat affected students differently in this study than in the Kellow and Jones (2005) study, the outcome was the same. White students scored statistically significantly higher than African American students when told that their test performance would be predictive of their performance on a statewide, high-stakes standardized test. In this respect, the implication is also the same: African American

students are at a disadvantage compared to White students when their knowledge and skills are measured using a high-stakes standardized test. These results suggest one reason that African American students are disproportionately left behind when they are retained in grade for failing to pass high-stakes tests. Unfortunately, their failure may be unrelated to their knowledge of the subject tested. As a result, African American students are at an unfair disadvantage compared with White students.

We want to make it clear that we are not suggesting that high-stakes tests are inherently biased in favor of White students. Rather, we believe that certain environmental and cognitive phenomena associated with high-stakes testing may serve to undermine the performance of African American students. Stereotypes will not disappear overnight; however, efforts can be made to minimize the impact of stereotypes on minority students facing a high-stakes test. Several promising studies have demonstrated the potential to reduce negative effects of stereotypes on the test performance of minority groups (e.g., Aronson, Fried, & Good, 2002; McIntyre, Paulson, & Lord, 2003).

Although this line of inquiry is, at best, in its nascent stages, researchers (Aronson et al., 2002) have suggested that changing students' views of intelligence from an entity to an incremental view may minimize the impact of stereotype threat. Students who believe that intelligence is fixed (entity view) are more likely to be less motivated and perform worse when faced with difficulties than students who believe that intelligence is changeable (incremental view). The reason for this is that students with an entity view are more likely to pursue performance goals (they want to demonstrate their intelligence and/or they prefer tasks that verify that they are smart and capable) than learning (mastery) goals (they are more concerned with learning new concepts and improving their competence). Researchers suggest that students in stereotype threat situations perceive themselves (maybe temporarily) like those who believe in a fixed intelligence. Therefore, in the same way that fixed intelligence leads to performance goals, we see that stereotype threat may lead to performance goals.

To combat these negative effects, researchers have suggested methods for changing students' beliefs about intelligence from entity to incremental. If students have an incremental view, they are more likely to adopt learning (mastery) goals and all of the positive outcomes associated with them. To change a student's belief from entity to incremental, researchers have tried a few approaches, such as showing students videos depicting the ways in which the brain changes every time something new is learned and having students write letters to (hypothetical) younger students reinforcing the view of intelligence as a malleable phenomenon (Aronson et al., 2002).

A second method of alleviating stereotype threat that shows promise is informing students about stereotype threat and its effects on students (Johns, Schmader, & Martens, 2005). This approach focuses on the attributions students make about the anxiety they may experience while taking challenging standardized tests. Students are encouraged to consider external, or situational attributions for their anxiety as a reaction to knowledge that negative societal stereotypes exist about them as a function of their group membership (e.g., African American). This stands in contrast to the tendency among many students to attribute their anxiety to internal, dispositional factors, such as being deficient in the domain being tested. Although studies have shown that anxiety is not reduced after exposure to information about the existence of stereotype threat, performance among members of stereotyped groups does improve, presumably because the new information provides a means for externalizing arousal.

Further investigation of programmatic interventions to address the influence of stereotypes on the success of African American and other minority students in high-stakes testing environments is clearly warranted. If indeed the "threat in the air," as Claude Steele (1997) so aptly stated, it is at least partially responsible for the depressed performance of stigmatized groups as a function of stereotypical thinking. As psychologists, we owe due diligence to develop methods to overcome this dilemma. Given the pervasive nature of stereotypical thinking, future research may address not just the psychological effects of these stereotypes on students but the attitudes and dispositions of teachers, school administrators, and parents.

REFERENCES

- Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.
- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*, 113-125.
- Aronson, J., & Steele, C. M. (2005). Stereotypes and the frailty of academic competence, motivation, and self-concept. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 436-456). New York: Guilford.
- 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 ed.). Cambridge, UK: Cambridge University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*, 1304-1312.

- Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5-18.
- Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist*, *41*, 1040-1048.
- Eccles, J. S. (1984a). Sex differences in achievement patterns. In T. Sonderegger (Ed.), *Nebraska symposium on motivation* (Vol. 32, pp. 97-132). Lincoln: University of Nebraska Press.
- Eccles, J. S. (1984b). Sex differences in mathematics participation. In M. Steinkamp & M. Maehr (Eds.), *Advances in motivation and achievement* (Vol. 2, pp. 93-137). Greenwich, CT: JAI.
- Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., et al. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75-146). San Francisco: Freeman.
- Eccles, J. S., Adler, T. F., & Meece, J. L. (1984). Sex differences in achievement: A test of alternate theories. *Journal of Personality and Social Psychology*, *46*, 26-43.
- Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin*, *21*(3), 215-225.
- Elliot, A., & Church, M. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, *72*, 218-232.
- Elliot, A. J., & McGregor, H. A. (2001). A 2 X 2 achievement goal framework. *Journal of Personality and Social Psychology*, *80*(3), 501-519.
- Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychological Science*, *16*(3), 175-179.
- Jones, G. M., Jones, B. D., & Hargrove, T. (2003). *The unintended consequences of high-stakes testing*. Lanham, MD: Rowman & Littlefield.
- Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles*, *47*(3/4), 193-198.
- Kellow, J. T., & Jones, B. D. (2005). Stereotype threat in African-American high school students: An initial investigation. *Current Issues in Education*, *8*(20). Retrieved October 25, 2005, from <http://cie.asu.edu/volume8/number20>
- Mayer, D. M., & Hanges, P. J. (2003). Understanding the stereotype threat effect with "culture-free" tests: An examination of its mediators and measurement. *Human Performance*, *16*, 207-230.
- McIntyre, R. B., Paulson, R. M., & Lord, C. G. (2003). Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology*, *39*, 83-90.
- Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its consequences for young adolescents' course enrollment intentions and performances in mathematics. *Journal of Educational Psychology*, *82*, 60-70.
- Middleton, M., & Midgley, C. (1997). Avoiding the demonstration of lack of ability: An underexplored aspect of goal theory. *Journal of Educational Psychology*, *89*, 710-718.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- Nicholls, J. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, *91*, 328-346.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

- Osborne, J. W. (2007). Linking stereotype threat and anxiety. *Education Psychology, 27*(1), 135-154.
- Pinel, E. C. (1999). Stigma consciousness: The psychological legacy of social stereotypes. *Journal of Personality and Social Psychology, 76*, 114-128.
- Pinel, E. C. (2002). Stigma consciousness in intergroup contexts: The power of conviction. *Journal of Experimental Social Psychology, 38*, 178-185.
- Pintrich, P. R. (1989). The dynamic interplay of student motivation and cognition in the college classroom. In C. Ames & M. L. Maehr (Eds.), *Advances in motivation and achievement: Motivation enhancing environments* (Vol. 6, pp. 117-160). Greenwich, CT: JAI.
- Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research, 31*, 459-470.
- Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications*. Upper Saddle River, NJ: Merrill Prentice Hall.
- Ployhart, R. E., Ziegert, J. C., & McFarland, L. A. (2003). Understanding racial differences on cognitive ability tests in selection contexts: An integration of stereotype threat and applicant reactions research. *Human Performance, 16*(3), 231-259.
- Ryan, K. E., & Ryan, A. M. (2005). Psychological processes underlying stereotype threat and standardized math test performance. *Educational Psychologist, 40*(1), 53-63.
- Schunk, D. H., & Pajares, F. (2005). Competence perceptions and academic functioning. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 141-163). New York: Guilford.
- Shell, D., Murphy, C., & Bruning, R. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology, 81*, 91-100.
- Smith, J. (2004). Understanding the process of stereotype threat: A review of mediational variables and new performance goal directions. *Educational Psychology Review, 16*(3), 177-206.
- Smithson, M. (2001). Correct confidence intervals for various regression effect sizes and parameters: The importance of non-central distributions in computing intervals. *Educational and Psychological Measurement, 61*, 605-632.
- Spencer, S., Steele, C. M., & Quinn, D. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4-28.
- Spielberger, C. D., Gorsuch, R. R., & Lushene, R. (1970). *The State-Trait Anxiety Inventory (STAI) test manual*. Palo Alto, CA: Consulting Psychologists Press.
- Stangor, C., Carr, C., & Kiang, L. (1998). Activating stereotypes undermines task performance expectations. *Journal of Personality and Social Psychology, 75*(5), 1191-1197.
- 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.
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 379-440). New York: Academic Press.
- Stricker, L. J. (1998). *Inquiring about examinees' ethnicity and sex: Effects on AP Calculus AB Examination performance* (College Board Rep. 98-1; ETS Research Rep. No. 98-5). New York: College Entrance Examination Board.
- Tobin, T. C., Ave, M., & Batts, D. W. (2004, May 16). Reading, writing, race. *St. Petersburg Times*, pp. A1, A8.
- Tobin, T. C., & Batts, D. W. (2004, May 17). In search of the causes. *St. Petersburg Times*, pp. A1, A9.
- Trochim, W. M. (2004). *The research methods knowledge base* (2nd ed). Chicago: Atomic Dog.

- U.S. Department of Education (2001). No Child Left Behind. Retrieved November 3, 2005, from <http://www.ed.gov/nclb/overview/intro/execsumm.html>
- Vartanian, L. R. (2000). Revisiting the imaginary audience and personal fable constructs of adolescent egocentrism. *Adolescence, 35*, 639-661.
- Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*, 456-457.
- Wicherts, J. M. (2005). Stereotype threat research and the assumptions underlying analysis of covariance. *American Psychologist, 60*, 267-268.
- Wiesen, J. P. (1996). *Spatial Ability Test*. Newton, MA: Applied Personnel Research.
- Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review, 12*, 265-310.
- Willson, V. L., & Kellow, J. T. (2002). Confusing the messenger with the message: A response to Bolon. *Education Policy Analysis Archives, 10*(9). Retrieved December 15, 2006, from <http://epaa.asu.edu/epaa>
- Wolters, C. A. (2004). Advancing achievement goal theory: Using goal structures and goal orientations to predict students' motivation, cognition, and achievement. *Journal of Educational Psychology, 96*(2), 236-250.