



Children's Stereotype Threat in African-American High School Students: An Initial Investigation

J. Thomas Kellow and Brett D. Jones
University of South Florida St. Petersburg

Stereotype threat refers to the risk associated with confirming a negative stereotype based on group membership. We examined this effect in a sample of African-American high school students. Stereotype threat was manipulated by presenting a visual spatial reasoning test as (a) diagnostic of mathematical ability or (b) a culture and gender fair test of mathematical reasoning. Support was found for the general effect, and while tests of the effect of the manipulation on anxiety and perceptions of ability and expectancies for success were statistically inconclusive, the data trended in the predicted direction. Implications related to the high-stakes testing of African-American students are discussed.

With the advent of the *No Child Left Behind* (NCLB) Act of 2001 (U.S. Department of Education, no date), the focus on high-stakes standardized testing has become even more intense as states have been coerced into developing testing programs to measure and report student achievement. To avoid leaving children behind, NCLB increased accountability for States, school districts, and schools and required students to be tested in grades three through eight in the basics of mathematics, reading or language arts, and science.

While the intent behind NCLB may be admirable, standardized testing is often problematic for minority students who, as a group, consistently score lower on standardized measures of achievement (Steele, 1997). In fact, African-American and Hispanic students continue to score well below White students on academic achievement tests (National Center for Education Statistics, 2004). This finding has often been labeled the "achievement gap" because of the large gap between the higher test scores of White students and the lower test scores of minority students. Although some progress has been

made (Coley, 2003), the achievement gap has failed to substantially close over several decades. As a result, researchers have searched for causes of the achievement gap, several of which have been identified, such as family and community differences (Jencks & Phillips, 1998), fewer opportunities for minority students to study a rigorous curriculum with highly qualified teachers (Oakes, Gamoran, & Page, 1992), social expectations (Lumsden, 1998; Rosenthal & Jacobson, 1968), and negative stereotypes concerning academic abilities (Steele, 1997).

To reduce the achievement gap, we suspect that some combination of these types of potential causes must be addressed. However, one of these potential causes, negative stereotypes concerning academic abilities, is of particular interest to us because so little research has been conducted about its effects on K-12 students. Steele and Aronson (1995) have researched this phenomenon with college students and labeled it "*stereotype threat*" (p. 797). Stereotype threat for an African-American student is the risk that she feels of confirming a

negative stereotype about her ethnic group. For instance, an African-American student might feel the risk of confirming the fact that African-American students generally score lower than Whites and Asians on standardized measures of achievement when she is taking a standardized mathematics test. In an era of high-stakes tests, stereotype threat is important to consider because it has been shown to have negative effects on student test scores (Steele & Aronson, 1995).

Although the effects of stereotype threat 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 K-12 students (Walton & Cohen, 2003). Given the unprecedented use of high-stakes testing in K-12 schools to determine grade promotion and graduation, this population seems deserving of study. If K-12 students experience stereotype threat during standardized state testing, this phenomenon could partially explain the achievement gap between White and minority students.

The purpose of this study was to investigate whether African-American high school freshman students experienced stereotype threat when given 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 exit-level criterion-referenced tests in various content domains in order 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. Second, these exit-level 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 grade level).

Narrative and Expository Schemata

Indeed, a large body of research shows that narrative and exposition do invoke different schemata. Traditional stories start by describing relations among settings, characters, and plots; they contain clear markers between episodes (e.g., beginnings and endings, setting and goal attainment); and conclude with a global ending (Mandler, 1984). Socio-linguistic models likewise find a well-defined structure in narrative discourse composed of six elements: abstract, orientation, complication, evaluation, result, and coda (Labov & Waletzky, 1967). The narrative structure of text has been shown to affect information reduction, organization, storage, and retrieval of textual information (Mandler & Johnson, 1977).

In contrast, expository discourse is organized into statements that allow readers to follow text flow through logic and causality, exposition relies primarily on declarative statements, logic, and reason, and it is evaluated in terms of its accuracy and strength of argument (Bruner, 1986, 1990). Expository text is generally analyzed in terms of propositional structure (e.g., Meyer & Rice, 1984; Otero & Kintsch, 1992). A well-written expository text uses a macroproposition early in the paragraph to organize subsequent propositions and connect them to the larger goals of the author.

Still, knowing how schemata differ does not tell us how these differences affect comprehension. This is not only dissatisfying from a research perspective, it creates practical limitations as well. There are many differences to be found between exposition and narrative, and therefore likely to be many sources of differences in processing. We focus on one that, based upon prior work in psycholinguistics, we believe to be of particular importance: That narrative and expository text tend to differ in the scope of processing they receive.

Literature Review

Stereotype Threat

The phenomenon of stereotype threat was first examined by Claude Steele and his colleagues at Stanford University (e.g., Steele & Aronson, 1995). Steele and Aronson proposed that differences in academic performance between minority and non-minority 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 related to group membership. Consistent with his hypothesis, Steele 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. Another popular manipulation of evaluation apprehension is to present a measure as a “traditional” test of achievement or intellect (evaluative) or as a “culture-free” or “non-biased” test (non-evaluative).

Interestingly, Steele (1997) found that, irrespective of presenting the task as evaluative or non-evaluative, simply indicating one’s race prior to taking the test was sufficient to activate stereotype threat. Similar findings have been produced between males and females using the same paradigm with mathematics performance as the dependent variable (e.g., Spencer, Steele, & Quinn, 1999). Moreover, the effect has been produced not only with academic achievement tests, but also with visual spatial

reasoning tasks (Mayer & Hanges, 2003; McKay, Doverspike, Bowen-Hilton, & Martin, 2002). A number of potential mediators of stereotype threat have been proposed. Of these, only anxiety and performance confidence have emerged as likely candidates (Smith, 2004).

Anxiety

A number of studies, beginning with Steele and Aronson (1995), have examined the role of anxiety in producing the stereotype threat effect (Smith, 2004). These researchers hypothesize that fear of confirming a negative stereotype related to one's group membership (e.g., African-American) elicits an anxiety response which, in turn, produces cognitive interference that undermines test performance. The anxiety variable 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 examined the role of anxiety in mediating stereotype threat are mixed, with some studies demonstrating an experimental effect on anxiety and others producing no effects.

In the present study, we chose to assess the role of anxiety because it potentially has 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.

Perceptions of Ability and Expectancy for Success

Eccles and her colleagues (Eccles et al., 1983; Eccles, Adler, & Meece, 1984; 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,b; 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).

Interestingly, the relationship between students' achievement level and their perceptions of ability and expectancies for success is not as strong for African-American students as it is for White students (Stevenson, Chen, & Uttal, 1990). This appears to be true, in part, because African-American students tend to have higher self-perceptions of ability and expectancies for success (Cooper & Dorr, 1995; Graham, 1994), yet have lower grades and lower levels of performance on standardized achievement tests (Graham, 1994). Stereotype threat researchers have speculated that facing a challenging test in a diagnostic context may lead minorities to question their ability (Steele and Aronson, 1995, call this an "ability-indicting" interpretation, p. 799) and perhaps withdraw effort to perform based on low expectations for success.

Rationale and Purpose of 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 one have used college students and other adult samples. The sole study conducted with high school students employed a very select sample (i.e., those taking the Advanced Placement Calculus Examination). Second, the majority of studies have used individual as opposed to group testing sessions. While this affords a certain amount of additional control over the experimental setting, it is hardly realistic in the context of real-world academic achievement testing. In the present study, we attempt to advance knowledge of stereotype threat by addressing these deficiencies in the literature while focusing on two possible mechanisms that may be associated with decrements in performance evinced by African-Americans facing a diagnostic challenge: (a) anxiety and (b) perceptions of ability and expectancies for success. We chose to use a visual spatial representation task in the experiment because it seemed better suited than a traditional achievement test when presented as a culture fair instrument (non-evaluative) or one highly correlated with academic achievement – in this case mathematics (evaluative).

Directional Hypotheses

H₁: After controlling for pre-existing differences in mathematics ability, African-American participants in the evaluative condition will score lower on the visual spatial task relative to White participants in the same condition;

H₂: After controlling for pre-existing differences in mathematics ability, 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;

H₃: African-American participants in the evaluative condition will report higher levels of anxiety than White participants in the same condition.

Method

Design

The present study employed a quasi-experimental 2 X 2 factorial design. The factors were ethnicity (African-American or White) and experimental condition (evaluative or non-evaluative). The dependent measures were (a) performance on a visual spatial reasoning task, (b) self-perceptions of ability and expectancies for success, and (c) level of state anxiety. Results from a standardized test of mathematics achievement were used as a covariate for analyses related to the first two measures.

Participants

Participants were recruited from freshman *critical thinking* courses at a large urban high school in Florida. Before initiating student recruitment, approval was obtained from the Institutional Review Board at the University of South Florida and the local school district. In addition, we received the approval and support of the principal at the targeted high school. All freshman students at this high school were required to enroll in this critical thinking course. Informed consent forms for parent or guardian approval were provided to 393 students. The study was presented as an investigation of mathematical reasoning. Parents were informed that the study had been approved by the school district and a university institutional review board. For their participation, students received a gift card for a free meal at a local restaurant. Ninety (23%) of the forms were signed and returned. The ethnic composition for students returning forms was: White 61%, African-American 36%, Hispanic 2%, and Asian 1%. Fifty-eight percent of the students were female. Based on the demographics of all 9th grade students at this school, Whites and females were slightly, but not statistically significantly, overrepresented in the sample. Since the focus of the study was on African-American and White students, data for the three Hispanic students and one Asian student that participated in the experiment were excluded from the final sample. In addition, five students (four White and one African-American) were absent on the day of the experiment, leaving a final sample of 81 participants.

Measures

After exposure to the evaluative or non-evaluative condition and completing the visual spatial task, all participants completed a seven-page survey.

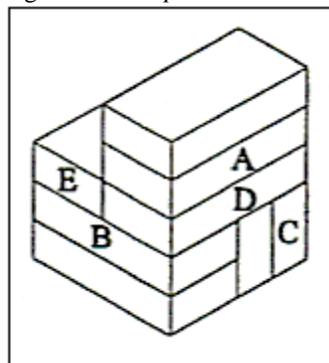
Survey items were ordered to address (a) perceptions of test bias, (b) state anxiety level, and (c) self-perceived ability and expectancies for success.

Test bias. A single item, "This test was biased against minority students," was used to assess the degree to which students felt the APR Visual Spatial Reasoning Task was a biased measure. Respondents indicated their belief using a seven-point Likert scale anchored at 1 (*not at all*) and 7 (*a lot*).

Florida Comprehensive Assessment Test (FCAT) in mathematics. The 8th grade criterion-referenced FCAT scale scores were used as a covariate in several analyses. The content of the test is linked to the Florida Sunshine State Standards, and has been thoroughly content validated. Concurrent validity coefficients based on a comparison with the Stanford Achievement Test (SAT-9) in mathematics are in the .75 to .85 range across grades. Internal consistency estimates using KR-20 range from .88 to .92.

APR Spatial Ability Test. 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).

Figure 1. Example APR test 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) geometry standards specifically state: "In grades 9-12 all students should visualize three-dimensional objects from different perspectives and analyze their cross sections" (NCTM, 2000, p. 308). The correlation of number

correct on the APR and FCAT mathematics scores was .4 in the present data. We performed a modified split-half procedure appropriate for speeded tests (Anastasi, 1988) on a sub-sample of 24 students in the present study and obtained a Spearman-Brown adjusted score reliability of .92. The APR served as the primary dependent variable in the study.

Self- and Task-Perception Questionnaire (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 five items from this scale that pertained to ability and performance expectations. Each item was scaled using a 7-point Likert format. 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, Adler, & Meece, 1984; Eccles & Wigfield, 1995) have empirically evaluated the psychometric properties of the scale and report strong factorial validity and reliability of scores. The correlation between the total score and FCAT mathematics in the present study was .4. We conducted an internal consistency analysis (Cronbach's alpha) on the present data and found a reliability estimate of .85.

State Trait Anxiety Inventory (STAI). A short-form of the STAI (Spielberger, Gorsuch, & Lushene, 1970) consisting of eight items with a four-point Likert scale anchored at 1 (*not at all*) and 4 (*very much*) was used to assess state anxiety of students while taking the APR test. Items were summed to produce a total score. A sample item is, "While taking the test I felt nervous." 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 ($r = -.15$). In addition, the scale yielded a respectable internal consistency estimate ($\alpha = .80$).

Procedure

The high school from which students were recruited was on a 4-block (period) schedule, so the four blocks (groups) of students were randomly assigned to be either (a) evaluative or (b) non-evaluative (there were no significant differences in prior mathematics ability between the four groups). Students were tested in groups during the period they ordinarily would have their critical thinking class. At each of the four blocks, students came from three different critical thinking classrooms to the school media center, which is often used for make-up testing on the FCAT (and thus, is not a novel testing

environment). The groups ranged in size from 14 to 24. There were a total of 46 students in the evaluative condition (28 White and 18 African-American) and 35 students in the non-evaluative condition (23 White and 12 African-American). After arriving at the media center, 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."

Non-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 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. After completing the example, students were 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."

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.

Results

Reporting and Analysis

For reporting purposes, the critical level for statistical significance (p) was set at .05. Exact p values for all results are reported, supplemented by partial eta-squared (η^2 part) as an uncorrected effect size measure. In addition, 95% confidence intervals (CI 95%) based on non-central F distributions were calculated for the effect size estimates using an SPSS routine developed by Smithson (2001). Overall model fit was assessed using unadjusted R^2 .

All analyses were conducted with SPSS 12.1 using the General Linear Model procedure. For all analyses, both main effects and interaction effects were examined. In addition, since the FCAT scores were statistically significantly related to both the APR ($r = .41$ and to the STPQ ($r = .38$), we included this measure as a covariate (after testing for violations of assumptions related to analysis of covariance) in these analyses.

Text Bias

The first analysis consisted of assessing the extent to which participants felt the APR Visual Spatial Reasoning Task was biased toward minorities. These data are found in **Table 1**.

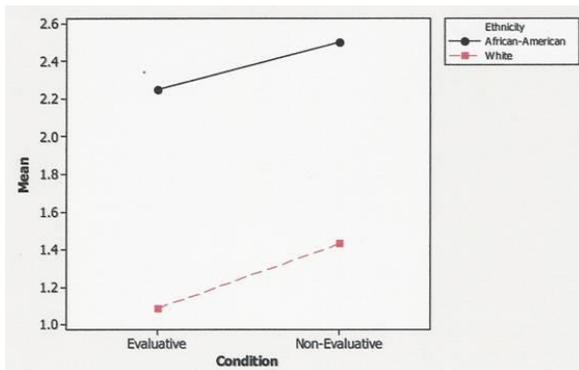
Table 1. Student perception of test bias

| EFFECT | F | p | Partial η^2 | 95% CI |
|-----------------------|------|-----|------------------|-----------|
| Ethnicity | 9.26 | .00 | .11 | .01 — .24 |
| Condition | 0.68 | .41 | .00 | .00 — .09 |
| Ethnicity * Condition | 0.02 | .90 | .00 | .00 — .01 |

Model $R^2 = .12$

There was a statistically significant effect for ethnicity that accounted for 11% of the variance in bias ratings. Neither condition nor the ethnicity by treatment interaction were statistically significant. The group means for this item are presented in Figure 2.

Figure 2. Mean perception of test bias against minority students by ethnicity and condition



African-American students were statistically significantly ($p = .00$) more likely to perceive the APR as biased, irrespective of experimental condition. Interestingly, Mayer and Hanges (2003) found a similar result using the Raven's Advanced Progressive Matrices, another visual spatial reasoning task. They attributed the lack of perceived bias in the evaluative group (as opposed to the non-evaluative condition) to the ambiguous nature of the test itself. Indeed, given the range of possible values (1 to 7) on this item, the overall means between evaluative ($X = 1.7$) and non-evaluative ($X = 2.0$) conditions suggest a relatively low presence of perceived test bias.

APR Visual Spatial Reasoning Task

Results of analysis of the primary dependent variable, the APR, are presented in **Table 2**. The FCAT scores in mathematics were used as a covariate in this analysis to control for pre-existing differences in achievement.

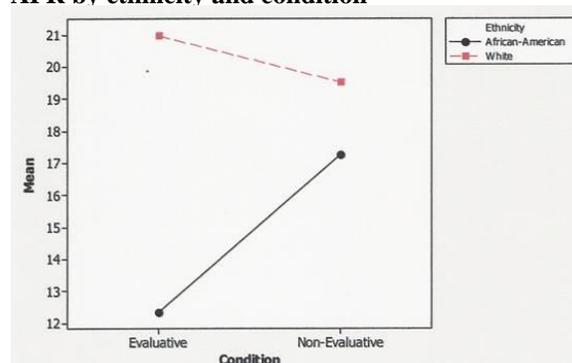
Table 2. Student performance on the APR by ethnicity and condition

| EFFECT | F | p | Partial η^2 | 95% CI |
|-----------------------|-------|-----|------------------|-----------|
| Ethnicity | 12.21 | .00 | .14 | .03 — .27 |
| Condition | 1.28 | .26 | .02 | .00 — .10 |
| Ethnicity * Condition | 4.06 | .04 | .05 | .01 — .16 |

Model $R^2 = .28$

A statistically significant main effect for Ethnicity ($p = .00$) was found that accounted for 14% of score variance. In addition, a statistically significant interaction effect ($p = .02$) for Ethnicity * Condition was present that captured 5% of score variance. The estimated marginal (adjusted) means are presented in Figure 3.

Figure 3. Mean adjusted number correct on the APR by ethnicity and condition



In the non-evaluative condition the performances of African-American ($X = 17.3$) and White students ($X = 19.5$) were similar. In the evaluative condition, however, White students ($X = 21.0$) outperformed

African-American students ($X = 12.3$) by nearly nine correct responses.

Self and Task Perception Questionnaire (STPQ)

Analysis of the STPQ data also employed the FCAT as a covariate. These results are found in Table 3.

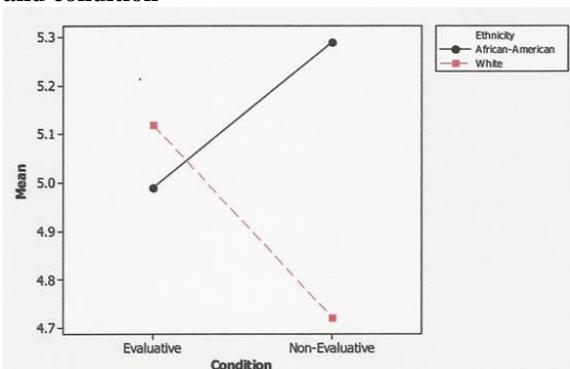
Table 3. STPQ scores by ethnicity and condition

| EFFECT | F | p | Partial η^2 | 95% CI |
|-----------------------|------|-----|------------------|-----------|
| Ethnicity | 0.61 | .44 | .01 | .00 — .08 |
| Condition | 0.03 | .86 | .00 | .00 — .02 |
| Ethnicity * Condition | 1.51 | .22 | .02 | .00 — .11 |

Model $R^2 = .19$

Neither the main effects nor interaction effect for the STPQ were statistically significant after statistically controlling for actual ability as measured by the FCAT mathematics scale. The plotted marginal means are displayed in Figure 4.

Figure 4. Mean adjusted STPQ scores by ethnicity and condition



Although not statistically significant the means trend in the predicted direction, with African-American students in the evaluative condition reporting slightly lower perceptions of mathematics ability and expectancies for success after controlling for actual ability.

State Trait Anxiety Inventory (STAI)

Results for the analysis of STAI scores are presented in Table 4.

Table 4. STAI scores by ethnicity and condition

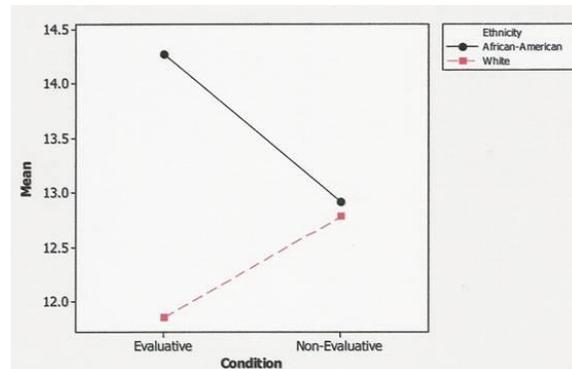
| EFFECT | F | p | Partial η^2 | 95% CI |
|-----------------------|------|-----|------------------|-----------|
| Ethnicity | 1.87 | .18 | .02 | .00 — .12 |
| Condition | 0.12 | .74 | .00 | .00 — .06 |
| Ethnicity * Condition | 1.49 | .23 | .02 | .00 — .11 |

Model $R^2 = .05$

Neither ethnicity nor condition was statistically significant. The hypothesized interaction effect was not statistically significant either, although there was

a modest η^2 part value of .02. The plotted means for the interaction may be found in Figure 5.

Figure 5. Mean STAI scores by ethnicity and condition

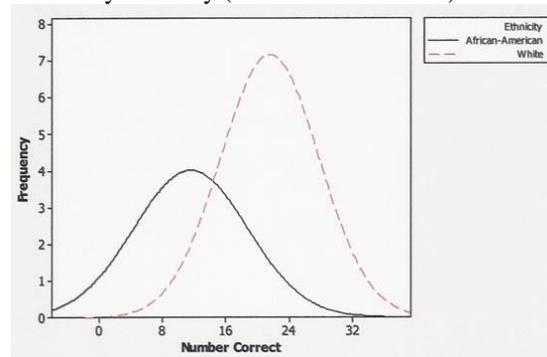


Despite the non-significant p value, the results are clearly in the hypothesized direction. Students as a whole in the non-evaluative condition were very similar in their self-reported levels of anxiety, while African-American students ($X = 14.3$) in the evaluative condition were more anxious than their White counterparts ($X = 11.9$).

Discussion

The present study represents the first effort of which we are aware to examine stereotype threat in an academically representative (as opposed to elite) high school African-American sample. The findings suggest that stereotype threat is indeed a phenomenon that exists with these students. African-American students in the evaluative condition scored significantly and substantially lower than White students on the APR. To put this effect into perspective, we calculated Cohen's d (Cohen, 1988) for the group mean differences using the pooled standard deviations. The effect size was 1.3, which is well above the 1.0 value Cohen suggested as a "large" effect. This effect size is illustrated in Figure 6, which shows the superimposed distributions of the APR results for the respective groups.

Figure 6. Distributional differences on APR number correct by ethnicity (evaluative condition)



Although there were no statistically significant differences between White and African-American students on the STPQ, the means trend in the directions we predicted. As mentioned previously, research generally indicates that African-Americans have higher self-perceptions of ability and expectancies for success (Cooper & Dorr, 1995; Graham, 1994). In the non-evaluative condition, African-Americans had higher perceptions of their competence. However, in the evaluative condition, their perceptions were slightly lower and very close to the mean values reported by Whites. If this trend were to hold true with a larger sample of students, it would indicate that the evaluative condition lowered African-American students' perceptions of their competence. Interestingly, it would also suggest that the evaluative condition slightly raised White students' perceptions of their competence. Walton and Cohen (2003) have labeled this phenomenon "stereotype lift" in non-minority students, whereby knowledge of the negative stereotype associated with minority students improves the performance, sense of competency, and social worth of the non-stereotyped group. Their meta-analysis of 43 stereotype threat studies provides some evidence for this effect.

Although the effect was not statistically significant and η^2 part was small, we did find that the interaction effect for the anxiety variable was in the expected direction. This provides at least some support for the potential role of anxiety as a concomitant of the stereotype threat effect. It also is interesting that White students in the evaluative condition tended to be less anxious, which again raises the possibility of a stereotype lift effect.

Limitations and Directions for Future Inquiry

In the present study we chose to use a visual spatial reasoning task rather than a traditional test of mathematics achievement. While we had legitimate reasons for selecting this option, one may wonder how the results might have differed if, instead, a traditional achievement test had been administered. Several recent studies (Mayer & Hanges, 2003; McKay, Doverspike, Bowen-Hilton, & Martin, 2002) have indicated that the stereotype threat effect is less pronounced when a visual spatial task, as opposed to a traditional achievement test, is used to assess performance. Thus, it is conceivable that a replication using a traditional test might yield even stronger effects than those reported in the present study.

Because of the relatively small ($n = 81$) size of the sample employed in the study we were limited in several ways. First, with only 30 African-American students, the cell sizes for this group in each condition were less than ideal. A larger sample would have improved the statistical power of the study (all other things being equal) and likely

allowed us to obtain statistically significant results for the last two tests. Second, we were unable to conduct formal tests of mediation using, for instance, a structural equation modeling approach with any real confidence. Third, Steele (1997) has suggested that the stereotype threat effect is most pronounced in individuals who are "identified" with the academic domain being tested. One way this variable has been operationalized is to study individuals who previously have demonstrated strong academic prowess in the domain as evidenced by high test scores (thus explaining his early focus on Stanford University students). In the present study we used previous mathematics achievement as a covariate in several analyses, which statistically adjusts the dependent variable scores to what they would have been if all participants were "equal" on the covariate. Unfortunately, they are not equated on any number of other moderating or mediating variables that are associated with academic achievement, so the approach is not entirely satisfactory.

In summary, while the present data suggest a rather strong stereotype threat effect in this population, the quest for definitive answers regarding the underlying causal mechanisms that produce the effect continues. Several researchers (e.g., Mayer & Hanges, 2003; Smith, 2004) have suggested that no one variable, such as anxiety, may explain the stereotype threat effect satisfactorily. Rather, the effect may be caused by a complex interaction between multiple mediating variables. Smith has suggested that other variables, such as achievement goals, play a part in the causal process. Future studies of this population might gain noteworthy insights into the causes of stereotype threat by employing larger samples and modeling multiple mediators and moderators simultaneously within a structural modeling framework. As Mayer and Hanges (2003) note, "It is possible that stereotype threat influences test scores through several mediators with only a slight affect on any one mediator. It is also possible that these variables influence one another making simple mediation models inappropriate" (p. 211). Finally, we are unaware of any systematic efforts to examine stereotype threat in Hispanic students, a rising demographic in the United States. It is reasonable to assume that these processes may be operating in other groups who are stereotyped based on low group performance. Indeed, during the debriefing process, two of the three Hispanic students whose data were excluded from the analysis pointedly asked the researchers, "What about Hispanic and multi-ethnic kids?"

Conclusion and Implications

The results of this study raise serious concerns about the use of high-stakes tests as a

measure of African-American high school students' knowledge and skills. These findings indicate that the achievement gap may be explained, in part, by the stereotype threat felt by African-American students during high-stakes tests. Consequently, compared to White students, African-American students may be at a disadvantage because they are unable to demonstrate their true abilities on tests in which African-Americans have been shown to consistently score lower than White students.

If stereotype threat and its associated mechanisms are indeed even partially responsible for decrements in the performance of African-American students, researchers and educators are morally and professionally obliged to uncover the specific mediators of this effect. Once these variables have been identified, it may be possible to develop effective interventions for minority students that would reduce the negative impact of stereotype threat. The implications of failing to address this charge, from a social justice perspective, are (at least) twofold and are embodied in the NCLB legislation. First, African-American children are disproportionately left behind by virtue of failing high-stakes tests and consequently being retained in grade. Second, schools with large African-American enrollments face sanctions based on the performance of these students. If tests are to be used in this fashion, we must at least have some assurance that the performance of African-American and other minority students is not being distorted by influences that are irrelevant to the task at hand.

References

- Anastasi, A. (1988). *Psychological Testing* (6 th ed.). New York: MacMillan.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2 nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Coley, J. (2003). *Growth in School Revisited: Achievement Gains from the Fourth to the Eight Grade*. Princeton, NJ: Educational Testing Service.
- Cooper, H. M., & Dorr, N. (1995). Race comparisons on need for achievement: A meta-analytic alternative to Graham's narrative review. *Review of Educational Research*, 65, 483-508.
- Eccles, J. S. (1984a). Sex differences in achievement patterns. In T. Sonderegger (Ed.), *Nebraska Symposium on Motivation* (Vol. 32, pp. 97-132). Lincoln, NE: Univ. 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 Press.
- Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75-146). San Francisco, CA: 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.
- Graham, S. (1994). Motivation in African Americans. *Review of Educational Research*, 64, 55-117.
- Jencks, C., & Phillips, M. (1998). *The Black-White Test Score Gap*. Washington, DC: Brookings Institution.
- Lumsden, L. (1998). Teacher expectations: What is professed is not always practiced. *Journal of Early Education and Family Review*, 5(3), 21-24.
- 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.
- McKay, P. F., Doverspike, D., Bowen-Hilton, D., & Martin, Q. D. (2002). Stereotype threat effects on the Raven Advanced Progressive Matrices scores of African Americans. *Journal of Applied Social Psychology*, 32, 767-787.
- 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.
- National Center for Educational Statistics (2004). *National assessment of educational progress: 2004 long-term trend assessment results*. Retrieved August 4, 2005, from <http://nces.ed.gov/nationsreportcard>
- National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.
- Oakes, J., Gamoran, A., & Page, R. N. (1992). Curriculum differentiation: Opportunities, outcomes, and meanings. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 570-608). Washington, DC: American Educational Research Association, 570-608.

- 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 Press.
- Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, 31, 459-470.
- Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom: Teacher expectation and pupils' intellectual development. New York: Rinehart and Winston.
- 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.
- 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.
- Stevenson, H., Chen, C., & Uttal, D. (1990). Beliefs and achievement: A study of black, white, and Hispanic children. *Child Development*, 61, 508-523.
- U.S. Department of Education (2001). *No Child Left Behind*. Retrieved December 15, 2003, from <http://www.ed.gov/nclb/landing.jhtml>
- 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.
- 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.
- .
- .
- .
- .

2005 Article Citation

Kellow T.J. & Jones, B.D. (2005, July 20). Stereotype Threat in African-American High School Students: An Initial Investigation. *Current Issues in Education* [On-line], 8(15). Available: <http://cie.ed.asu.edu/volume8/number20/>

Author Notes

J. Thomas Kellow
College of Education
University of South Florida St. Petersburg
140 Seventh Ave South, COQ 201, St. Petersburg, Florida 33701
kellow@stpt.usf.edu

J. Thomas Kellow is an Assistant Professor of Measurement and Research at the University of South Florida St. Petersburg. His research interests include high-stakes testing, applied statistics, and program evaluation methodology. He received his Ph.D. in Educational Psychology from Texas A&M University-College Station.

Brett D. Jones.
College of Education
University of South Florida St. Petersburg
bjones@stpt.usf.edu

Brett D. Jones is an Assistant Professor of Educational Psychology. His professional interests include applying cognitive and motivational theories to instruction and examining the effects of high-stakes testing on students, teachers, and administrators. He has published several articles related to test-based accountability, as well as a book entitled *The Unintended Consequences of High-Stakes Testing*.

Note:

This work was supported in full or in part by a grant from the University of South Florida St. Petersburg New Investigator Research Grants Fund. This support does not necessarily imply endorsement of the research conclusions by the University.

Note from the 2015 Executive Editor, Constantin Schreiber

August 8, 2015. This article was first published at the original *Current Issues in Education* website, located at <http://cie.asu.edu/articles/index.html>. In 2009, *CIE* changed online platforms to deliver the journal at <http://cie.asu.edu>. The original *CIE* website was from then on only used as an archival repository for published articles prior to Volume 12. After the new *CIE* website moved to a different server in 2014, the original website and original article URLs could not be accessed anymore. Therefore, this article had to be repurposed into the published format you are viewing now.

All content from the original publication has been preserved. No content edits occurred. Spelling, grammar, and mechanical errors that may be found were present in the original publication. The *CIE* logo and publisher information in use at the time of the article's original publication is unaltered. Please direct questions about this article's repurposing to cie@asu.edu.

2015 Article Citation

Kellow T. J., & Jones, B. D. (2005). Stereotype threat in African-American high school students: An initial investigation. *Current Issues in Education*, 8(20). Retrieved from <http://cie.asu.edu/ojs/index.php/cieatasu/article/view/1663>



Current Issues in Education

Mary Lou Fulton College of Education
Arizona State University