



ELSEVIER

Applied Developmental Psychology 24 (2003) 645–662

**Applied
Developmental
Psychology**

Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat

Catherine Good^{a,*}, Joshua Aronson^{b,1}, Michael Inzlicht^b

^a*Department of Psychology, Columbia University, 405 Schermerhorn Hall, 1190 Amsterdam Avenue,
New York, NY 10027, USA*

^b*New York University, East Building, 239 Greene Street, 537F, New York, NY 10003, USA*

Abstract

Standardized tests continue to generate gender and race gaps in achievement despite decades of national attention. Research on “stereotype threat” (Steele & Aronson, 1995) suggests that these gaps may be partly due to stereotypes that impugn the math abilities of females and the intellectual abilities of Black, Hispanic, and low-income students. A field experiment was performed to test methods of helping female, minority, and low-income adolescents overcome the anxiety-inducing effects of stereotype threat and, consequently, improve their standardized test scores. Specifically, seventh-grade students in the experimental conditions were mentored by college students who encouraged them either to view intelligence as malleable or to attribute academic difficulties in the seventh grade to the novelty of the educational setting. Results showed that females in both experimental conditions earned significantly higher math standardized test scores than females in the control condition. Similarly, the students—who were largely minority and low-income adolescents—in the experimental conditions earned significantly higher reading standardized test scores than students in the control condition.

© 2003 Published by Elsevier Inc.

Keywords: Stereotype threat; Adolescents; Standardized tests; Mathematics; Reading; Gender differences; Low-income students; Minority students; Attributions; Beliefs about intelligence

* Corresponding author.

E-mail addresses: cgood@psych.columbia.edu (C. Good), joshua.aronson@nyu.edu (J. Aronson), michael.inzlicht@nyu.edu (M. Inzlicht).

¹ Tel.: +1-212-998-5543; fax: +1-212-995-4563.

1. Introduction

When Neil de Grasse Tyson received his doctoral degree from Columbia University in 1991 he became the seventh African American astrophysicist out of 4000 astrophysicists nationwide. In his convocation address delivered at Columbia's graduation ceremonies, he summarized his life path by noting:

In the perception of society my athletic talents are genetic; I am a likely mugger/rapist; my academic failures are expected; and my academic successes are attributed to others. To spend most of my life fighting these attitudes levies an emotional tax that is a form of intellectual emasculation (de Grasse Tyson, 1991).

The “emotional tax” de Grasse Tyson referred to aptly describes the undermining role that stereotypes can play in the intellectual lives of African American students—and, indeed, of anyone who contends with stereotypes that question his or her abilities. The psychological literature has taught us much about how negative stereotypes can contribute to intellectual underperformance among stigmatized students (e.g., Steele, 1997; Steele, Spencer, & Aronson, 2002). It has offered much less, however, about what can be done to lift the emotional tax levied upon stigmatized students. In this article, we hope to address this shortcoming by introducing an intervention program designed to remedy stereotype-based underperformance at a critical period in a student's burgeoning academic life—the transition to junior high school.

1.1. *The academic performance of minorities and females*

The academic underperformance of Black and Hispanic students and the underrepresentation of girls and women in the mathematics and science domains have long troubled people concerned with educational inequities. Each year, statistics from statewide and national tests reaffirm the disturbing pattern of underachievement. For example, compared to White and Asian students, Black students receive lower grades and have higher dropout rates at practically every level of schooling. Yet despite decades of national concern, recent data suggest that the disparities are not likely to disappear soon. Indeed, the gap in high school GPA between Black and White students actually increased in 2002 (The College Board, 2002), and Black students still obtain lower scores on standardized tests of reading, math, and science (Jencks & Phillips, 1998). Because standardized test scores are the preferred standard for college admissions, it is not surprising that Black students make up less than 10% of admissions to 4-year colleges (National Center for Education Statistics, 2003). Although Hispanic students fare somewhat better, their school achievement and standardized test scores also tend to lag substantially behind that of White and Asian students (see Romo & Falbo, 1995). A similar pattern exists for girls and young women in the areas of math and science. For example, females lag behind boys on the math section of the SAT by as much as 35 points. This performance gap has decreased by a mere 3 points in the past 10 years (The College Board, 2002) despite the numerous programs designed to increase females' math and science outcomes, such as *Expanding Your Horizons* (<http://www.expandingyourhorizons.org/>).

Much of the psychological and educational research examining the various factors presumed to underlie these race and gender gaps have concluded that sociological factors, such as teachers' expectations, are often to blame (e.g., Eccles & Jacobs, 1992; Jencks & Phillips, 1998; Romo & Falbo, 1995; Sadker & Sadker, 1994; Valencia, 1997). Recent research in social psychology, however, has demonstrated that these gaps may be a product of a more general cognitive process that may be, as a result, more amenable to intervention than previously thought (Aronson et al., 1999; Aronson, Quinn, & Spencer, 1998; Spencer, Steele, & Quinn, 1999; Steele, 1997; Steele & Aronson, 1995; Steele et al., 2002). Specifically, this research suggests that individuals may suffer negative performance outcomes (lower standardized test scores and less engagement with academics) because they are burdened by the prospect of confirming cultural stereotypes impugning their intellectual and academic abilities. Calling this burden "stereotype threat," Steele and Aronson (1995) conducted research demonstrating that this phenomenon can be felt as a physiological arousal (Ben-Zeev, Fein, & Inzlicht, 2003; Blascovich, Spencer, Quinn, & Steele, 2001) that often results in substantial decrements in intellectual performance. For example, Steele and Aronson showed that inducing stereotype threat—by presenting a test as a measure of intellectual ability, or by asking test takers to indicate their race before the test—can significantly undermine African Americans' performance on intellectual tasks. They also showed that reducing stereotype threat—by convincing test takers that the test is not being used to measure their abilities—can significantly improve African Americans' performance, dramatically narrowing the race gap. Building on these findings, subsequent experimental work has established that stereotype threat can undermine the academic performance of females in math (Inzlicht & Ben-Zeev, 2000; Spencer et al., 1999; Good, Aronson, & Harder, 1999), students from low socioeconomic backgrounds (Croizet & Claire, 1998) and, in fact, any group that contends with negative stereotypes about their intellectual abilities (Aronson et al., 1999). This research suggests that underperformance results, in part, from pejorative interpretations of failure, which are facilitated by the stereotype (Aronson, Fried, & Good, 2002). These pejorative interpretations, which suggest low ability rather than surmountable challenges, add stress and self-doubt to students' educational experiences and diminish their sense of belonging to the academic arena (Good & Dweck, 2003).

Two factors that contribute to stereotype threat can help us determine the most effective strategies for helping academically stigmatized individuals overcome vulnerability to stereotype-based underperformance. First, research indicates that evaluative scrutiny is at the heart of most situations that evoke stereotype threat. Being evaluated in a stereotyped domain is sufficient to trigger the trademark responses associated with stereotype threat—lack of enjoyment of the educational process, increased anxiety and stress, and, ultimately, underperformance. Second, group composition—the racial or gender mix in a room of test takers—also can trigger stereotype-relevant thoughts, and thus vulnerability to stereotype threat (Inzlicht & Ben Zeev, 2000, *in press*) because group composition can make salient one's social identity and the stereotypes associated with that identity.

Clearly, these two conditions—evaluative scrutiny and identity salience—are characteristic of most testing environments in which students find themselves, such as Advanced Placement examinations, the SAT, the GRE, and the like. Group salience is endemic to these testing

situations, not only because students often indicate their race and gender before taking the test, but also because minority and female students take the tests in the presence of White students and males. And, undoubtedly, students fully recognize that their performance on these tests can have important implications for their academic futures, determining college credit, scholarships, and school admissions. Consequently, arming students with the means to overcome the stereotype threat they are likely to experience during these tests could potentially reduce the race and gender gaps that have troubled standardized testing for decades.

In designing the intervention program for this study, we were driven by two goals. The first was to develop a program for adolescents, for whom educational difficulties can develop quickly and can set the stage for future academic problems. The second goal was to explore the potential of social psychological interventions for improving standardized test performance. Such interventions have successfully boosted test performance in the laboratory (Aronson et al., 1998) and have improved GPAs (Aronson et al., 2002; Wilson & Linville, 1985). However, this past research involved college undergraduates and did not involve actual high-stakes standardized testing.

1.2. Stereotype threat and the transition to junior high school

Our desire to implement a program for adolescents led us to focus on the transition to junior high school. Timing our intervention to coincide with adolescents' transition to junior high school is important for three reasons. First, the transition to junior high school is the time at which most students falter academically, some continuing to struggle throughout their academic life (Eccles, Lord, & Midgley, 1991). Sadly, the decline is particularly steep and there is less likely to be a rebound for girls in math and for minority students more generally. For example, although most students experience some initial difficulties transitioning to a new school environment, several studies report significantly more problems, i.e., suspensions, low academic performance, conflicts with parents, etc., among Black and Latino students than among White students (e.g., Felner, Primavera, & Cauce, 1981; Simmons, Black, & Zhou, 1991). Furthermore, it is during junior high school that early differences in confidence manifest as differences in math performance between males and females. One possible reason for this is the "stereotype climate" that is engendered and reinforced by the middle school setting (Aronson & Patnoe, 1997).

Second, the developmental literature suggests that it is not until adolescence that stereotyped students should regularly experience the stress and underperformance that accompanies explicit evaluations of ability (see Aronson & Good, 2003, for a review). Supporting this prediction, recent stereotype threat research has shown that most children are not meaningfully affected by stereotype threat until the age of 11 or 12 (Good & Aronson, 2003; McKown & Weinstein, 2003).

Third, junior high school teachers themselves may unwittingly exacerbate students' vulnerability to stereotype threat. For example, junior high teachers are more likely than teachers of younger children to believe that their students' abilities are fixed and less likely to believe that students can increase their abilities through instruction (see Anderman et al.,

2001). Moreover, as Aronson et al. (2002) argue, anything that promotes a fixed ability mindset exacerbates stereotype threat (also see Aronson, 2002). Consequently, adolescence, particularly the time when students transition to junior high school, may be the most appropriate and effective time to intervene and help students cope with societal stereotypes.

Our second goal was to identify interventions that have had particular success in increasing students' academic outcomes, for both stereotyped and nonstereotyped groups. In designing an effective intervention that would combat vulnerability to stereotype threat, we turned to the social psychology literature.

1.3. Addressing attributions: Effective interventions

As discussed previously, stereotype threat disrupts academic performance because the stereotypes provide a pejorative explanation for struggle and difficulty. That is, they raise the possibility—at least in the mind of stereotyped individuals—that the academic difficulties they experience may be due to an internal fault or shortcoming, namely, that they lack the ability to succeed on the task. In a landmark study, Wilson and Linville (1985) designed a clever intervention that addressed precisely these types of pejorative explanations for negative outcomes.

Wilson and Linville (1985) argued that these pejorative explanations produce a vicious cycle that leads to and perpetuates poor performance. That is, self-blame for a negative outcome leads to increased anxiety, in turn resulting in poor performance and even more self-blame. Wilson and Linville suggested that to stop the cycle it can help to shift the blame from pejorative attributions (one's lack of intelligence) to nonpejorative ones (the difficulty of the context), and that this might halt the downward spiral of anxiety and poor performance and ultimately lead to improved academic outcomes. To test this hypothesis, Wilson and Linville convinced entering college students to “reattribute” their academic difficulties from stable internal causes to temporary, external causes. Specifically, they taught the participants in their study that not only do most entering freshmen experience academic difficulties, but that these difficulties lessen after the first year (unstable cause) and are most likely due to the difficulties inherent in transitioning from high school to college (external cause). To make their argument convincing, the participants were exposed to statistics documenting the fact that many first-year college students actually improve their GPAs after the first year. They were also shown videotapes of upperclassmen discussing their college experiences. In these videotapes, the upperclassmen reported that their GPAs were markedly lower in their first year of college.

The results were remarkable. The intervention participants not only improved their grades in the second year of college, but also were more likely than nonintervention participants to remain in college. Because of these findings, we hypothesized that changing stereotyped students' explanations for difficulty from pejorative to nonpejorative can help reduce vulnerability to stereotype threat. In particular, we predicted that stereotyped students who were exposed to a mentoring program that focused on changing attributions for difficulty from pejorative to nonpejorative would outperform students who were exposed to a mentoring program that did not include a focus on attributions.

In a conceptually similar intervention, Aronson et al. (2002) addressed the implicit beliefs that students have about intelligence in an effort to improve the academic performance of minority college students. Decades of research have shown that students who think of intellectual ability as a fixed trait (entity theory) rather than as a potential that can be developed (incremental theory) are at greater risk of negative academic outcomes—decreased confidence, loss of enjoyment, and performance impairment—when faced with difficulties or setbacks (Dweck & Sorich, 1999; Henderson & Dweck, 1990; Jourden, Bandura, & Banfield, 1991; Martocchio, 1994; Tabernero & Wood, 1999; Wood & Bandura, 1989). Noting that stereotype threat elicits many of the hallmark responses characteristic of entity theorists, Aronson et al. hypothesized that because stereotypes imply fixed, limited ability based on group membership, stereotype threat may temporarily induce an entity-theory mind-set. Consequently, stereotype threat could be overcome by adopting an incremental-theory mind-set (Aronson et al., 2002).

In their intervention, Aronson et al. (2002) induced intervention participants to adopt an incremental mind-set by having them watch a highly compelling film depicting the ways the brain changes every time something new is learned. To reinforce this message, the students also participated in an ostensible pen pal program in which they wrote a letter to a struggling junior high student. In their letter, they emphasized the idea that intelligence is expandable and increases with mental work. Results showed that students who had training in the incremental theory reported greater enjoyment of their academic work and greater valuing of academics in general than students in the control group who did not receive the incremental training. In addition, they showed a clear gain in GPA over the other groups. Other interventions in malleability training have resulted in similar academic gains (e.g., Blackwell, Dweck, & Trzesniewski, 2003). Because of these findings, we hypothesized that encouraging stereotyped adolescents to view intelligence as malleable can help reduce vulnerability to stereotype threat. In particular, we predicted that stereotyped students who were exposed to a mentoring program that encouraged an incremental view of intelligence would outperform students who were exposed to a mentoring program that did not include a focus on the malleability of intelligence.

1.4. Goals and hypotheses

The results of these interventions for students' academic achievement are encouraging—they all increase important student school outcomes, such as grades and academic enjoyment. Although grades are certainly important, we wondered whether we could use such techniques to reduce the effects of stereotype threat—and thus raise performance—on standardized tests, which despite continued controversy are becoming more prevalent as markers of learning and merit. Thus, our intervention program tested the efficacy of these past interventions on junior high school students' performance on the state-administered standardized test. We hypothesized that implementing an intervention that addressed either the pejorative explanations for academic difficulty, or the maladaptive beliefs about the nature of intelligence that stereotypes foster, could reduce the gaps in standardized test performance that typically widen when students enter junior high school (Eccles et al., 1991). Specifically, we predicted that the math performance on the statewide test would be higher for females trained to make nonpejorative

attributions for difficulty or to appreciate the malleability of intelligence than for females who did not receive this training. We made parallel predictions for the reading test, namely, that all students given either of the two interventions would outperform students in a control group.

2. Method

2.1. Overview

In the present study, we investigated whether teaching junior high school students about different perspectives on school achievement could reduce their vulnerability to stereotype threat and increase their standardized test performance. To test our hypotheses, we designed an in-depth intervention in which we taught seventh graders messages we hypothesized would help them cope with stereotype threat. Specifically, we taught seventh-grade students one of two educational messages, or a combination of both messages that we thought would reduce the effects of stereotype threat on their standardized test performance. These messages were conveyed to the seventh graders by college students who mentored the junior high students throughout the school year. All of the students in the study received a college student mentor with whom they discussed various issues, including adjusting to the new school environment and useful study strategies. For one group of students, the mentors also discussed the expandable nature of intelligence and helped students learn more about how the brain is able to form new connections throughout one's lifetime. For a second group of students, the mentors explained that all students face academic difficulty during the transition to junior high school, but that over time, most students are able to overcome these difficulties and reach high levels of achievement. For a third group of students, the mentors combined these two messages. To reinforce and help students internalize the messages, the students created web pages that advocated, in their own words and pictures, the experimental messages that the students were learning from their mentors. The standardized test performance of these three groups was compared to that of a control group of students who was mentored about the dangers of drug use and who created antidrug web pages. In this way, we were able to determine whether the positive outcomes we were predicting resulted from adopting one (or both) of the two messages about achievement, rather than mere participation in the mentoring program.

We predicted that relative to the control group, the participants in the three experimental conditions would receive higher scores on the statewide standardized tests of math and reading, which they took at the end of the school year. Furthermore, we predicted that the benefits of this intervention would occur primarily for the Hispanic students on the reading test, and the female students on the math test.

2.2. Participants and design

This study was conducted in a rural school district in Texas that served a largely low-income population, comprised of 63% Hispanic, 15% Black, and 22% White population. Of the Hispanic students, 92% came from homes where the parents spoke English fluently.

Approximately 70% of these students were qualified to receive reduced cost or free lunches. Of the participants in our study, 67% were Hispanic, 13% were Black, and 20% were White. In addition, 45% of the students were female and 55% were male. Given the research showing stereotype threat effects for students who are Black (Steele & Aronson, 1995), Hispanic (Aronson & Salinas, 1999), female (Spencer et al., 1999), and low-income (Croizet & Claire, 1998), all of the participants in the sample were potentially susceptible to stereotype threat.

A total of 138 seventh-grade students (both male and female) who were enrolled in a computer skills class as part of their junior high curriculum participated in the study. Enrollment in the course was randomly determined by the school administration and all students in the course participated in the study. As part of the course curriculum, students learned a variety of computer skills, including e-mailing and web page design. Shortly after the school year began (mid-October), students in the class were randomly assigned a mentor, with whom they communicated in person and via e-mail throughout the school year. Students also were randomly assigned to one of the four experimental conditions that determined which educational message they learned: incremental, attribution, combination, and antidrug conditions. In the incremental condition, students learned about the expandable nature of intelligence; in the attribution condition, students learned about the tendency for all students to initially experience difficulty during seventh grade but then to experience improvement; in the combination condition, students learned about the expandable nature of intelligence and about the temporarily difficult transition to seventh grade; in the antidrug condition (the control condition), students learned about the perils of drug use. At the end of the year, students took statewide standardized tests in math and reading.

2.3. Procedure

2.3.1. Mentors

At the beginning of the fall semester (early September) approximately 25 college student mentors from the University of Texas participated in a 3-h training session in which they completed a required mentor-training course designed by the school district. As a supplement to the district's course, the mentors learned methods of conveying each of the four experimental messages. In order to reduce the possibility of demand characteristics, we told the mentors that we predicted students in all conditions of the experiment to benefit academically from the mentoring program, but the mentors remained blind to the specific hypotheses of the study. To avoid possible effects due to individual differences between mentors, we required each mentor to work with students in each of the four experimental conditions. However, because of scheduling restrictions with the junior high school, this was not possible. Instead, each mentor was randomly assigned mentoring responsibilities for one to two students in three of the four experimental conditions. Thus, each mentor worked with approximately six students.

2.3.2. Computer course

As mentioned previously, the participants in the study were seventh-grade students who were enrolled in a computer skills course at the junior high school. At the beginning of the

school year, students learned basic keyboarding skills such as typing, mouse skills (drag, drop, click, etc.), and Internet navigation. As the students became proficient in these rudimentary skills, they proceeded to more advanced computer topics such as e-mailing and web page design. As a final project in the course, students were required to design their own web page. The computer course instructor told the students that they would each receive a college student mentor who would be available to help them design and build their web page. The instructor further told the students that although the mentors would not be able to come to the class each week, they would be available to answer any questions the students had about their web page or about any problems they were having with school through e-mail.

2.3.3. *Intervention procedure*

After students became proficient in the basics of computer use and e-mailing (mid-November), we began the intervention. The mentors met with their students in person for 90 min in mid-November, and then again for 90 min at the beginning of the second semester (end of January). All remaining communication occurred via the Internet through an e-mail program created specifically for this study.

The mentors served three purposes. First, they provided useful advice for the students regarding study skills and any adjustment problems the students may have experienced during the difficult transition to junior high school. Second, they explicitly taught the students one of the four experimental messages: The expandable nature of intelligence (incremental condition), the tendency for all students to initially experience difficulty but then bounce back (attribution condition), a combination of these two messages (combined condition), or the perils of drug use (antidrug control condition). The mentors conveyed these messages in person during the two school visits and through weekly e-mail correspondence with the students throughout the school year. Finally, the mentors helped the students design and create a web page in which the students advocated, in their own words and pictures, the experimental message conveyed by the mentor throughout the year. The mentors told the students that their web pages would serve as public service announcements for other students who were having difficulty in school. Research has repeatedly demonstrated that such advocacies are extremely effective means of getting individuals to adopt the beliefs they are induced to advocate (e.g., Higgins & Rholes, 1978).

To further help students internalize the message, and to give them ideas about what to put on their own web page, we designed a “restricted web space” for each of the experimental conditions. In these web spaces, students could “surf” the restricted web and learn in more detail the experimental message. For each condition, access to the restricted web was limited to the condition in which they were assigned. In other words, students in one condition could access the restricted web space that provided information relevant to their condition but not the other conditions. Moreover, while surfing the restricted web, the students could not inadvertently access web pages not affiliated with the study. The mentors encouraged the students in each condition to use the most convincing elements from these restricted web spaces in their own web page.

2.3.3.1. Incremental condition. Participants in this condition learned that intelligence is not a finite endowment, but rather an expandable capacity that increases with mental work. To reinforce the scientific validity of this perspective, the mentors taught students some facts about the brain and how it works. For example, students learned about the role of neurons and dendrites and how the brain is capable of forming new neural connections throughout one's life. In addition to hearing this perspective directly from the mentors, students also explored the restricted web space to learn in more detail how the brain works. For example, numerous web pages within the restricted web incorporated animated pictures of the brain, scientific images of neurons and dendrites, and narrative explanations to demonstrate how the brain forms new neural connections when it is engaged in effortful problem solving. Other pages contained various testimonies and catch phrases regarding the expandability of intelligence, such as "The mind is a muscle; the more you use it, the stronger it grows."

2.3.3.2. Attribution condition. Participants in this condition learned that many students tend to experience difficulty when they move to a new educational situation (such as junior high school) but then bounce back after they become accustomed to their new environment. The mentors explained that in response to academic setbacks, many students erroneously conclude that they are not capable of high academic achievement when, in fact, the difficulties they experience are more likely due to the novelty of the situation. To illustrate, the mentors described their own difficulty in making the adjustment to junior high but that they eventually overcame these difficulties. Furthermore, the mentors pointed out that many aspects of junior high are very different from elementary school, such as changing classes each period, attempting more difficult subjects, meeting many more students, and adjusting to many teachers' teaching styles rather than just one or two. In short, the mentors encouraged the students to shift their attributions for difficulties from pejorative to nonpejorative causes, that is, from their own shortcomings to the novelty of the situation.

As in the incremental condition, students in the attribution condition were encouraged to explore a restricted web space containing information that reinforced this perspective. For example, the students viewed bar graphs showing the average school grades for seventh and eighth graders. The graphs demonstrated that most students earn poor grades in seventh grade—a C average—but that by eighth grade, students perform much better, earning an A average. Other pages contained similar graphs showing that enjoyment of school and school attendance both improve from seventh to eighth grade. Additionally, students viewed other pages that contained testimonials from older students proclaiming that the difficulties of seventh grade get easier over time.

2.3.3.3. Combined condition. Participants in this condition learned both the incremental message and the attribution message. They also explored both restricted web spaces.

2.3.3.4. Antidrug control condition. Participants in this condition learned about the perils of drug use. Specifically, they learned that in addition to the health consequences of using drugs, drugs could also interfere with academic achievement. To reinforce this message, students

explored a restricted web space to learn about the harmful effects of a variety of drugs, such as amphetamines, cocaine, marijuana, alcohol, and tobacco.

2.3.4. *Dependent measure*

At the end of the school year, we analyzed students' math and reading achievement as measured by the Texas Assessment of Academic Skills (TAAS) test, a statewide standardized achievement test administered to all students in the district. Many school districts in Texas use the test scores to determine whether a student will be promoted to the next grade or will be retained in the current grade. For example, students who fail to earn a standard score of at least 70 out of a maximum of 100 are often retained in the current grade or are not given course credit for a particular subject. Moreover, it is customary for schools in Texas, as elsewhere, to be evaluated and rewarded as a function of their aggregate test scores. For example, students' standardized test scores are used to determine if a school is meeting minimum educational requirements for their students. Principals of schools that have a high proportion of students failing to achieve the minimum standard score of 70 are often dismissed. Thus, students, teachers, administrators and parents consider the test to be high stakes, and it is therefore the kind of test likely to produce stereotype threat (Aronson et al., 1999).

3. Results

3.1. *Outlier analysis*

For the math test and the reading test, outliers were identified using the first and third quartiles (Q1 and Q3) and the interquartile range (IQR). Specifically, scores that were less than $Q1 - 1.5(IQR)$ or greater than $Q3 + 1.5(IQR)$ were not included in the analyses (Iglewicz & Hoaglin, 1993). Because we did not use prior ability as an inclusion criterion for the study, and because some of the students in the study spoke limited English, it was necessary to conduct an outlier analysis. For the analysis of the math scores, this criterion resulted in eliminating five participants' scores from the 138 participants who took the math test. For the analysis of the reading scores, this criterion resulted in eliminating six participants' scores from the 135 who took the reading test. It is important to note that the eliminated students did not tend to come from any condition in particular.

3.2. *Math achievement test performance*

To determine if the intervention influenced students' math standardized test scores, math TAAS scores were submitted to a 2 (gender) \times 4 (experimental condition) analysis of variance (ANOVA).² The ANOVA yielded a significant main effect for condition, $F(3,125)$

² Without removing the outliers, the ANOVA resulted in a significant main effect of condition, $F(3,137) = 2.98$, $p = .03$, qualified by a marginally significant interaction between sex and condition, $F(3,137) = 2.52$, $p = .06$.

= 7.24, $p = .001$, and a significant main effect for gender, $F(1,125) = 4.30$, $p = .04$. These main effects were qualified by a significant gender by condition interaction, $F(3,125) = 2.98$, $p = .03$ (see Fig. 1). Planned comparisons indicated that the males in the antidrug condition ($M = 81.55$, $SD = 6.03$) performed significantly better on the math test than the females ($M = 74.00$, $SD = 8.37$), $t(30) = 3.21$, $p = .002$, Cohen's $d = 1.04$. However, in all other conditions, the gender gap in math performance disappeared (all ps *ns*). Importantly, the large effect size between males and females that is present in the control condition was completely erased in all three experimental conditions. Furthermore, although the three experimental manipulations (incremental, attribution, and combined) each increased both male and female students' math scores as compared to the control condition (antidrug), the manipulations appeared to be particularly beneficial for the female students, as one would expect if their performance were being suppressed by stereotype threat. Compared to the control condition, females achieved significantly higher math scores if they were in the incremental condition ($M = 82.11$, $SD = 5.72$), $t(26) = 3.34$, $p = .001$, Cohen's $d = 1.13$; the attribution condition ($M = 84.53$, $SD = 5.41$), $t(25) = 4.29$, $p = .001$, Cohen's $d = 1.50$; or the combined condition ($M = 84.06$, $SD = 7.09$), $t(26) = 4.14$, $p = .001$, Cohen's $d = 1.30$. Again, these are all large effect sizes, indicating that the intervention procedures meaningfully increased females' math scores compared to the control condition. The only marginally significant differences on math scores for male students in the four conditions occurred between the control (antidrug) condition ($M = 81.55$, $SD = 6.03$) and the incremental condition ($M = 85.25$, $SD = 5.42$), $t(40) = 1.95$, $p = .054$, Cohen's $d = .64$ (all other ps *ns*).

3.3. Reading achievement test scores

To determine if the intervention influenced students' reading test scores, reading TAAS scores were submitted to a one-way ANOVA comparing performance of students in the four

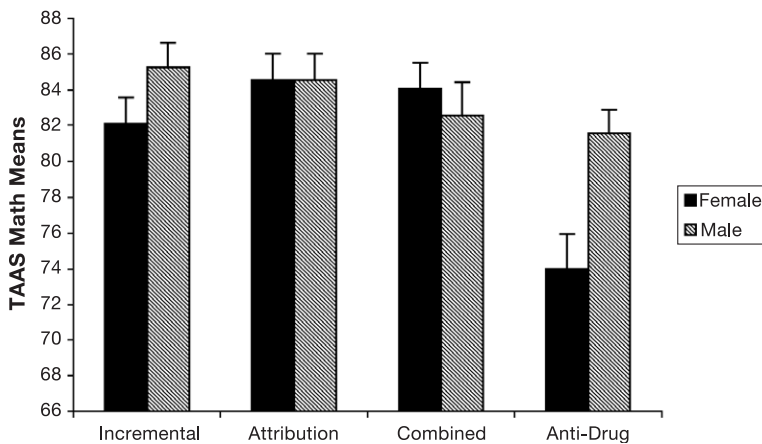


Fig. 1. Average math scores on the TAAS test.

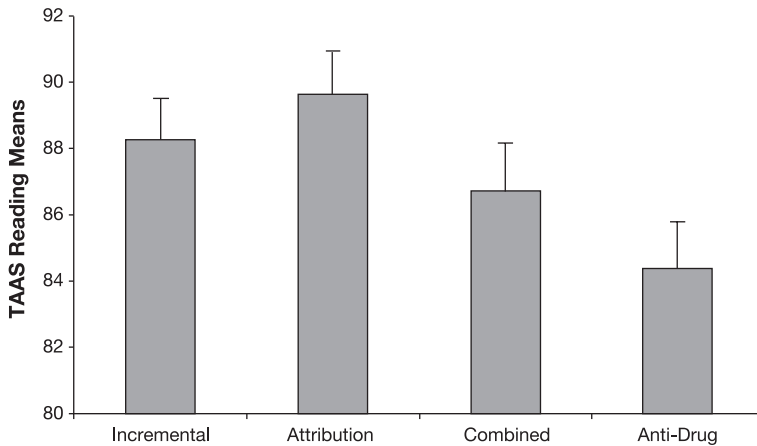


Fig. 2. Average reading scores on the TAAS test.

experimental conditions.³ Although we were interested in comparing reading achievement for minority and White students, the sample did not contain enough White students to perform these comparisons. The ANOVA revealed a significant effect of condition, $F(3,125) = 2.71$, $p = .05$ (see Fig. 2). Planned comparisons indicated that students in the incremental condition ($M = 88.26$, $SD = 7.17$) and students in the Attribution condition ($M = 89.62$, $SD = 7.01$) achieved significantly higher scores on the reading test than students in the control condition ($M = 84.38$, $SD = 7.79$), $t(65) = 2.07$, $p = .041$, Cohen's $d = .52$; and $t(61) = 2.72$, $p = .008$, Cohen's $d = .71$, respectively. There were no differences between the combined condition ($M = 86.71$, $SD = 8.70$) and the other conditions.

4. Discussion

Findings were consistent with hypotheses for the most part. Results showed that the typical gender gap in math standardized test performance emerged for the participants in the control condition. That is, boys outperformed girls on the math test if they had been mentored in the harmful consequences of drug use. However, when the participants learned about the expandability of intelligence the gender gap in math performance disappeared. The incremental condition increased both boys' and girls' math performance, but this increase in math scores was particularly pronounced for the female students, which is consistent with predictions derived from analysis of stereotype threat processes. Similarly, the gender gap in math performance disappeared when participants were encouraged to make nonpejorative attributions for their difficulties and when they were exposed to both the incremental and the reattribution intervention message.

³ Without removing the outliers, the ANOVA did not result in a significant main effect of condition. Post hoc analyses, however, confirmed that students in the incremental and attribution conditions scored significantly higher on the reading test than did the students in the control condition.

As with math performance, a similar pattern of increased achievement was found on the reading test. Students who were mentored in the malleability of intelligence performed better on the reading test than students who were mentored in the perils of drug use. Because the participants in the study were primarily from low socioeconomic backgrounds and 80% of the participants were either Black or Hispanic, all were potentially vulnerable to experiencing stereotype threat (Aronson & Salinas, 1999; Croizet & Claire, 1998; Steele & Aronson, 1995). Consequently, encouraging adolescents to make nonpejorative explanations for difficulty—that is, to think of intelligence as expandable rather than fixed or to attribute difficulties to the novelty of the situation rather than their own shortcomings—can meaningfully increase student achievement, especially for those students who face negative stereotypes about their abilities such as Hispanic, Black, low-income, and female students.

The results of this intervention were encouraging. Stereotyped students—females in math, and ability-stigmatized students in reading—increased their standardized test scores after participating in the intervention program. Initially, we hoped to determine which of the intervention messages—the incremental nature of intelligence or the reattribution training—would have the most beneficial effects. Furthermore, we had originally hypothesized that combining the incremental message with the reattribution message would increase students' standardized test scores more than either message alone. Interestingly, we found that both messages increased students' standardized test scores. Combining the two messages did not appear to have an additive effect, which was surprising since the two intervention messages were, at least superficially, distinct. In hindsight, however, the lack of an additive effect is understandable.

Although the two intervention messages implemented in this intervention appear different on the surface, they are at heart, very similar. Importantly, closer inspection of Dweck's work on implicit theories of intelligence reveals a clear kinship to attribution theory. As Dweck and her colleagues repeatedly have shown, the attributions one makes for poor performance depend upon one's beliefs about the nature of intelligence (see Dweck, 1999, for a review). This research has shown, for example, that entity theorists are more likely to blame their own shortcomings for academic difficulties or failures. That is, they make internal, stable attributions for negative outcomes. Not coincidentally, these are precisely the kinds of pejorative attributions that attribution theory predicts can lead to the downward spiral of self-blame, anxiety, and underperformance. Alternatively, incremental theorists are much less likely to blame their own intellectual shortcomings for their current struggles. Instead, incremental theorists view mistakes as an indicator that they did not try hard enough or did not approach the problem appropriately. In other words, they make external, unstable attributions for negative outcomes. Clearly, encouraging students to view intelligence as expandable does not simply change their beliefs about intelligence; more importantly, it also changes the attributions they make for the causes of their difficulties. Given the underlying similarities of the constructs, it is not surprising, then, that both techniques led to higher test scores and did not produce additive effects when both were presented in combination. It is quite plausible that both interventions simply addressed the same underlying concerns.

Regardless of the specific message that led to the improvements, we note the ease with which our intervention led to significant increases in students' standardized test scores.

Student performance did not improve through additional skills drilling or cramming of content related to the test. Rather, students improved by learning attitudes that helped them contend with the anxieties that research has shown develop in part from their social identities. This, we believe, is important not only as validation of a theoretically derived psychological intervention, but also as grounds for further questioning the use of standardized tests as markers of ability and learning. Moreover, this adds a somewhat discouraging note to our findings. Standardized test scores may be poor predictors of a student's future academic success (e.g., Jencks & Phillips, 1998), yet they are nevertheless used as indicators of a student's current and future abilities. Students with high SAT scores have a much better chance of being admitted to the college of their choice. Similarly, statewide standardized test scores in elementary and junior high school are no longer simply used as indicators of a school's effectiveness; they increasingly are used to determine whether or not a student may proceed to the next grade, earn credit in a required course, and ultimately, graduate from high school.

Our intervention significantly boosted the performance of girls, minority, and low-income students by addressing the psychologically threatening nature of these assessments. Hence, our results provide further evidence of the fragility of the standard measures used to assess learning, potential, and ultimately, success. Despite this sobering note, it is gratifying to see performance rise in response to a psychological intervention. Indeed, widespread endorsement of the belief in malleable intelligence and nonpejorative attributions for difficulties may render standardized testing more equitable for students who must contend with stereotypes impugning their intellectual abilities.

Based on the previous discussion one might wonder whether our intervention succeeded because it directly reduced or eliminated stereotype threat altogether. We believe that despite the success of our intervention, stereotype threat remains a potential obstacle for students to overcome. As Aronson et al. (2002) demonstrated, encouraging students to adopt an incremental view of intelligence does not alleviate perceptions of a stereotype-threatening environment. In their study, African American participants reported greater feelings of being judged by others through the lens of stereotypes than did White students, regardless of experimental condition. Thus, as in the Aronson et al. study, our intervention likely succeeded by changing stereotyped students' responses to a stereotype threatening situation rather than changing their direct perception of stereotype threat.

Clearly, many junior high students—particularly Black, Hispanic, low-income, and female students—may experience the intellectual emasculation that Dr. de Grasse Tyson felt during much of his academic life. The current research is encouraging because it demonstrates a successful strategy for stemming the spiral of self-blame, anxiety, and underperformance that many adolescents experience.

Acknowledgements

This research was supported by a William T. Grant Scholars award and a grant from the Russel Sage Foundation to Joshua Aronson, and by a Spencer Foundation/American

Educational Research Association fellowship awarded to Catherine Good. We also thank the Center for Research on Culture, Development, and Education for supporting this work.

We would like to thank Jeannetta Williams and David Disko for their assistance in conducting this research. We also thank Karen Rester for her helpful comments on an earlier draft.

References

- Anderman, E. M., Eccles, J. S., Yoon, K. S., Roeser, R., Wigfield, A., & Blumenfeld, P. (2001). Learning to value mathematics and reading: Relations to mastery and performance-oriented instructional practices. *Contemporary Educational Psychology*, *26*, 76–95.
- Aronson, E., & Patnoe, S. (1997). *The jigsaw classroom*. New York: Longman.
- Aronson, J. (2002). Stereotype threat: Contending and coping with unnerving expectations. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 279–301). San Diego, CA: Academic Press.
- Aronson, J., Fried, C., & 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., & Good, C. (2003). The development and consequences of stereotype vulnerability in adolescents. In F. Pajares, & T. Urdan (Eds.), *Adolescence and education: Vol. 2. Academic motivation of adolescents* (pp. 299–330). Greenwich, CT: Information Age Publishing.
- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, *35*, 29–46.
- Aronson, J., Quinn, D., & Spencer, S. (1998). Stereotype threat and the academic performance of minorities and women. In J. Swim, & C. Stangor (Eds.), *Prejudice: The target's perspective* (pp. 83–103). San Diego, CA: Academic Press.
- Aronson, J., & Salinas, M. (1999). *Stereotype threat: The role of effort withdrawal and apprehension on the intellectual underperformance of Mexican-Americans*. Unpublished manuscript.
- Ben-Zeev, T., Fein, S., & Inzlicht, M. (in press). Arousal and stereotype threat. *Journal of Experimental Social Psychology*.
- Blackwell, L. S., Dweck, C. S., & Trzesniewski, K. (2003). *Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention*. (Unpublished manuscript). New York: Columbia University.
- Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science*, *12*, 225–229.
- The College Board (2002). *2002 College bound seniors: A profile of SAT program test takers* [On-line]. Available: http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2002/pdf/2002_TOTAL_GROUP_REPORT.pdf. Retrieved May 20, 2003.
- Croizet, J., & Claire, T. (1998). Extending the concept of stereotype and threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, *24*(6), 588–594.
- de Grasse Tyson, N. (1991). *PhD convocation address: St Paul's Chapel, Columbia University* [On-line]. Available: <http://research.amnh.org/users/tyson/speeches/PhDConvocationAddress.html>. Retrieved January 25, 2003.
- Dweck, C. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia: Psychology Press.
- Dweck, C. S., & Sorich, L. (1999). Mastery-oriented thinking. In C. R. Snyder (Ed.), *Coping* (pp. 232–251). New York: Oxford University Press.

- Eccles, J. S., & Jacobs, J. E. (1992). The impact of mothers' gender-role stereotypic beliefs on mothers' and children's ability perceptions. *Journal of Personality and Social Psychology*, *63*, 932–944.
- Eccles, J. S., Lord, S., & Midgley, C. (1991). What are we doing to early adolescents? The impact of educational contexts on early adolescents. *American Journal of Education*, *99*, 521–542.
- Expanding Your Horizons. Available: <http://www.expandingyourhorizons.org/>. Retrieved April 16, 2003.
- Felner, R. D., Primavera, J., & Cauce, A. M. (1981). The impact of school transitions: A focus for preventive efforts. *American Journal of Community Psychology*, *9*, 449–459.
- Good, C., & Aronson, J. (2003). *The development of stereotype threat and its relation to theories of intelligence: Effects on elementary school girls' mathematics achievement and task choices* (Unpublished manuscript). New York: Columbia University.
- Good, C., & Dweck, C. S. (2003). *The effects of a stereotype threatening environment on women's sense of belonging to mathematics* (Unpublished raw data). New York: Columbia University.
- Good, C., Aronson, J., & Harder, J. A. (1999). *Stereotype threat in the absence of a kernel of truth: Women in Calculus*. (Unpublished manuscript). Austin: The University of Texas.
- Henderson, V., & Dweck, C. S. (1990). Achievement and motivation in adolescence: A new model and data. In S. Feldman, & G. Elliott (Eds.), *At the threshold: The developing adolescent* (pp. 308–329). Cambridge, MA: Harvard University Press.
- Higgins, E. T., & Rholes, W. S. (1978). Saying is believing: Effects of message modification on memory and liking for the person described. *Journal of Experimental Social Psychology*, *14*, 363–378.
- Iglewicz, B., & Hoaglin, D. C. (1993). *How to detect and handle outliers*. Milwaukee, WI: American Society for Quality Control.
- Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*, 365–371.
- Inzlicht, M., & Ben-Zeev, T. (in press). Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing. *Journal of Educational Psychology*.
- Jencks, C., & Phillips, M. (1998). *The Black–White test score gap*. Washington, DC: Brookings Institution Press.
- Jourden, F., Bandura, A., & Banfield, J. (1991). The impact of conceptions of ability on self-regulatory factors and motor skill acquisition. *Journal of Sport and Exercise Psychology*, *13*, 213–226.
- Martocchio, J. J. (1994). Effects of concepts of ability on anxiety, self-efficacy, and learning in training. *Journal of Applied Psychology*, *79*, 819–825.
- McKown, C., & Weinstein, R. (2003). The development and consequences of stereotype-consciousness in childhood. *Child Development*, *74*, 498–515.
- National Center for Education Statistics (2003). *The Condition of Education* (Publication No. 2003-067). Available: NCES Web site <http://nces.ed.gov/pubs2003/2003067.pdf>. Retrieved June 20, 2003.
- Romo, H., & Falbo, T. (1995). *Latino high school graduation: Defying the odds*. Austin, TX: University of Texas Press.
- Sadker, M., & Sadker, D. (1994). *Failing at fairness: How our schools cheat girls*. New York: Touchstone.
- Simmons, R. G., Black, A., & Zhou, Y. (1991). African-American versus white children and the transition to junior high school. *American Journal of Education*, *99*, 481–520.
- Spencer, S., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*, 4–28.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology*, *69*, 797–811.
- Steele, C. M., Spencer, S., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. Zanna (Ed.), *Advances in experimental social psychology* (pp. 379–440). San Diego, CA: Academic Press.
- Tabernero, C., & Wood, R. E. (1999). Implicit theories versus the social construal of ability in self-regulation and performance on a complex task. *Organizational Behavior and Human Decision Processes*, *78*, 104–127.

- Valencia, R. R. (1997). Latinos and education: An overview of sociodemographic characteristics and schooling conditions and outcomes. In M. Barrera-Yepes (Ed.), *Latino education issues: Conference proceedings* (pp. 13–37). Princeton, NJ: Educational Testing Service.
- Wilson, T. D., & Linville, P. W. (1985). Improving the performance of college freshmen with attributional techniques. *Journal of Personality and Social Psychology*, *49*, 287–293.
- Wood, R., & Bandura, A. (1989). Impact of conceptions of ability on self-regulatory mechanisms and complex decision making. *Journal of Personality and Social Psychology*, *56*, 407–415.