



Reducing stereotype threat in classrooms: a review of social-psychological intervention studies on improving the achievement of Black students



**NATIONAL CENTER** FOR  
**EDUCATION EVALUATION**  
AND **REGIONAL ASSISTANCE**

Institute of Education Sciences  
U.S. Department of Education



# Reducing stereotype threat in classrooms: a review of social-psychological intervention studies on improving the achievement of Black students

July 2009

**Prepared by**

**Joshua Aronson**  
Department of Applied Psychology  
New York University

**Geoffrey Cohen**  
Department of Psychology  
University of Colorado at Boulder

**Wendy McColskey**  
SERVE Center at  
University of North Carolina at Greensboro

**With contributions from**

**Bianca Montrosse**  
SERVE Center at  
University of North Carolina at Greensboro

**Karla Lewis**  
SERVE Center at  
University of North Carolina at Greensboro

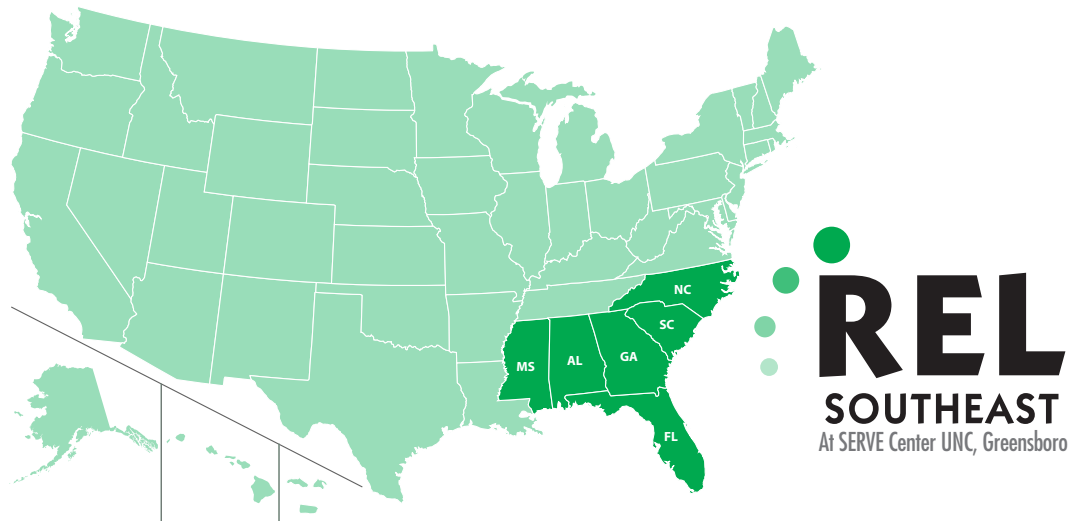
**Kathleen Mooney**  
SERVE Center at  
University of North Carolina at Greensboro



**NATIONAL CENTER** FOR  
**EDUCATION EVALUATION**  
AND **REGIONAL ASSISTANCE**

Institute of Education Sciences

U.S. Department of Education



**Issues & Answers** is an ongoing series of reports from short-term Fast Response Projects conducted by the regional educational laboratories on current education issues of importance at local, state, and regional levels. Fast Response Project topics change to reflect new issues, as identified through lab outreach and requests for assistance from policymakers and educators at state and local levels and from communities, businesses, parents, families, and youth. All Issues & Answers reports meet Institute of Education Sciences standards for scientifically valid research.

July 2009

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-06-CO-0028 by Regional Educational Laboratory Southeast administered by the SERVE Center at the University of North Carolina at Greensboro. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Aronson, J., Cohen, G., McColskey, W., Montrosse, B., Lewis, K., and Mooney, K. (2009). *Reducing stereotype threat in classrooms: a review of social-psychological intervention studies on improving the achievement of Black students* (Issues & Answers Report, REL 2009–No. 076). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from <http://ies.ed.gov/ncee/edlabs>.

This report is available on the regional educational laboratory web site at <http://ies.ed.gov/ncee/edlabs>.

# Reducing stereotype threat in classrooms: a review of social-psychological intervention studies on improving the achievement of Black students

**Stereotype threat arises from a fear among members of a group of reinforcing negative stereotypes about the intellectual ability of the group. The report identifies three randomized controlled trial studies that use classroom-based strategies to reduce stereotype threat and improve the academic performance of Black students, narrowing their achievement gap with White students.**

This review located and summarized the findings of randomized controlled trial studies on classroom-based social-psychological interventions aimed at reducing the experience of stereotype threat that might otherwise lead some Black students to underperform on difficult academic tasks or tests. Reducing the achievement gap between Black and White students is a critical goal for states, districts, and schools. Experimental research on both inducing and reducing stereotype threat can inform discussions of strategies.

Some students may perform below their potential because of the stress of being under constant evaluation in the classroom. Black students, however, may experience another source of stress in addition to this general one (which they share with their nonminority

peers). This second source of stress is specific to negatively stereotyped groups. It arises from a fear of reinforcing negative stereotypes about the intellectual ability of their racial group. Because Black students must contend with two sources of stress rather than one, their performance may be suppressed relative to that of their nonminority peers.

A systematic search was conducted for empirical studies of classroom-based social-psychological interventions designed to reduce stereotype threat and thus improve the academic performance of Black students. Search term combinations, such as “stereotype threat” and “intervention,” and “achievement gap” and “intervention,” were used to search a number of bibliographic databases. In addition, a web site on this topic with an extensive reference list was also reviewed. This initial search identified 289 references. After applying relevant inclusion criteria for topical and sample relevance, three experimental studies were identified. The three studies found positive impacts on the academic performance of Black students for the following social-psychological strategies:

- Reinforce for students the idea that intelligence is expandable and, like a muscle, grows stronger when worked.

- Teach students that their difficulties in school are often part of a normal learning curve or adjustment process, rather than something unique to them or their racial group.
- Help students reflect on other values in their lives beyond school that are sources of self-worth for them.

These three experiments are not an exhaustive list of the interventions to consider in reducing the racial achievement gap, nor are they silver bullets for improving the academic

performance of Black students. Rather, they present scientific evidence suggesting that such strategies might reduce the level of social-psychological threat that some Black students might otherwise feel in academic performance situations. It is important to note that while the strategies use established procedures that can be emulated by teachers and administrators, they also require thought and care on the part of schools and teachers in applying them in their particular situations.

**July 2009**

---

**TABLE OF CONTENTS**

<b>Why this study?</b>	<b>1</b>
Need for the study	1
What is stereotype threat and how has it been studied?	2
<b>Findings of three experimental studies of interventions to reduce stereotype threat in grade 7 classroom settings</b>	<b>4</b>
Study 1: Blackwell, Trzesniewski, and Dweck 2007, “Implicit theories of intelligence predict achievement across an adolescent transition: a longitudinal study and an intervention”	4
Study 2: Good, Aronson, and Inzlicht 2003, “Improving adolescents’ standardized test performance: an intervention to reduce the effects of stereotype threat”	7
Study 3: Cohen, Garcia, Apfel, and Master 2006, “Reducing the racial achievement gap: a social-psychological intervention”	9
<b>Concluding thoughts on turning research into practice</b>	<b>12</b>
<b>Notes</b>	<b>16</b>
<b>Appendix A Research on the relationship between stereotype threat and Black students’ academic performance</b>	<b>17</b>
<b>Appendix B Methodology</b>	<b>20</b>
<b>Appendix C Article screening and study quality review protocols</b>	<b>31</b>
<b>References</b>	<b>35</b>
<b>Boxes</b>	
1 Study methods	3
2 Key terms	6
<b>Figures</b>	
1 Estimated mean math scores by experimental condition	7
2 Percentage of Black and White students receiving a grade of D or lower in targeted course in same semester as the intervention, by experimental condition	12
<b>B1 First- and second-level screening and assessment of the quality of studies</b>	<b>21</b>
<b>Tables</b>	
1 Reported intervention impacts on the spring grade 7 state reading test (Texas Assessment of Academic Skills)	9
2 Covariate-adjusted mean grade point average (averaged over both studies) for intervention and control groups, by level of preintervention performance	11
3 Summary of effects reported by the three studies	15
<b>B1 Search results</b>	<b>20</b>
<b>B2 Disposition of references</b>	<b>22</b>
<b>B3 Quality of final studies included in report</b>	<b>26</b>
<b>B4 Methodological summary of the three experimental studies reviewed</b>	<b>28</b>

**Stereotype threat arises from a fear among members of a group of reinforcing negative stereotypes about the intellectual ability of the group. The report identifies three randomized controlled trial studies that use classroom-based strategies to reduce stereotype threat and improve the academic performance of Black students, narrowing their achievement gap with White students.**

## WHY THIS STUDY?

At every level of family income and school preparation, Black students<sup>1</sup> on average earn relatively lower grade point averages (GPAs) and scores on standardized tests (Bowen and Bok 1998; Hacker 1995; Jencks and Phillips 1998; Steele 1997). In a society where economic opportunity depends heavily on scholastic success, even a partial narrowing of the achievement gap would lead to a positive change in the lives of many academically at-risk children.

### Need for the study

Regional Educational Laboratory Southeast serves six southeastern states for which reducing the achievement gap between Black students and White students continues to be a major concern. The data indicate an education crisis in the Southeast Region, especially for Black male students (KewalRamani et al. 2007; Wald and Losen 2005). A report by the Southern Regional Education Board (SREB) on SAT and ACT scores concludes that between 1998 and 2002 none of the 16 SREB states narrowed the achievement gap between Black and White students (Southern Regional Education Board 2003). The achievement gap even widened for Black male students. Among the SREB states, which include the six states covered by the Regional Educational Laboratory Southeast, only 45 percent of Black male students graduated from high school in 2003 compared with 61 percent of Black female students, 65 percent of White male students, and 67 percent of White female students.

Thus, Regional Educational Laboratory Southeast frequently receives requests from Southeast Region educators for information on new ideas on interventions, programs, and policies that could close the achievement gap between Black and White students. Several Southeast Region states have regularly hosted conferences on this topic and published reports based on reviews.

Many potential contributing factors in the achievement gap have been explored, some

**Because of an awareness of negative stereotypes presupposing academic inferiority, Black and other minority students may worry that they could confirm the intellectual inferiority alleged by such stereotypes. Such worries can hinder their test performance, motivation, and learning**

individual (for example, family socioeconomic background, self-efficacy, and student aspirations) and some school related (for example, class size, distribution of resources across schools, quality and diversity of teachers, and lack of explicit and high performance standards). Other factors external to school are also suggested as having an impact on racial gaps in academic performance, such as a lack of high-quality early childhood education and of economic

opportunities to pursue postsecondary education, an important incentive to do well in school. (For reviews of research on the achievement gap, see Bowen and Bok 1998; Jencks and Phillips 1998; Rothstein 2002.) This report recognizes that there is a complex set of influential factors and that many of them are beyond a teacher's influence; these are not addressed here. Rather, to respond to the ongoing need for new information in this area, this review located and summarized findings from experimental studies on classroom-based social-psychological interventions to reduce stereotype threat in schools and classrooms that might lead some Black students to underperform on difficult academic tasks or tests.<sup>2</sup>

### What is stereotype threat and how has it been studied?

What is stereotype threat? Social psychologists hypothesize that racial stigma could help explain why, on average, Black and White students of similar socioeconomic backgrounds perform differently in college and on key standardized tests (Steele and Aronson 1995; see also Steele 1997). As students progress through school, classroom learning environments may become increasingly competitive, evaluative in nature, and stressful for some minority students. The logic behind stereotype threat is that because of an awareness of negative stereotypes presupposing academic inferiority, Black and other minority students may worry that they could confirm the intellectual inferiority alleged by such stereotypes (see appendix

A for a summary of the research on stereotype threat). Such worries, in turn, can hinder their test performance, motivation, and learning.

Research on stereotype threat began with laboratory studies exploring why Black college students seemed to be performing below their potential. Although a test-taking situation may seem objectively the same for all students, some students, because of their social identity, may experience it in a very different way. Steele and Aronson (1995) conducted a seminal experiment to explore the negative impact of administering a test under potentially stereotype-threat-inducing conditions by randomly assigning study participants to two different test-taking conditions. In one test-taking condition, a standardized test (composed of verbal Graduate Record Exam items) was presented to one group of college students as “diagnostic of intellectual ability.” It was hypothesized that Black students in this condition would worry that performing poorly could confirm a stereotype about their racial group's intellectual ability. Black students performed worse in this condition than when the same test was given in a second condition that introduced the test as one that was “not diagnostic of your ability.” The two ways of introducing the test had no effect on the performance of White students. Black students in the study sample answered roughly 8 of 30 test items correctly in the “threat” condition and roughly 12 of 30 correctly in the “no threat” condition.

Since the original experimental studies on the effects of inducing stereotype threat (Steele and Aronson 1995; Steele 1997), there has been an explosion of research documenting the negative effect of this phenomenon on performance of various types (for reviews see Ryan and Ryan 2005; Shapiro and Neuberg 2007; Steele, Spencer, and Aronson 2002; Walton and Cohen 2003; Wheeler and Petty 2001). Shapiro and Neuberg (2007, p. 125), in reviewing this literature, suggest that

The intellectual excitement surrounding the stereotype threat concept and research program stems in large part from the possibility



## BOX 1

**Study methods**

*Search and screening.* This study began with a thorough search, screening, and quality review to identify empirical studies of classroom-based social-psychological interventions designed to reduce stereotype threat and thus improve the academic performance of Black students. In addition to literature searches using key terms, a web site on this topic with an extensive reference list of peer-reviewed journal articles was examined ([www.reducingstereotypethreat.org](http://www.reducingstereotypethreat.org)). The literature search yielded 158 citations, and the web site reference list yielded an additional 131 citations.

The 289 references were then screened for inclusion using a set of six questions (see appendix C for the article screening protocol). A total of 214 studies were excluded based on the initial screening, applying the first three criteria (see table B2 and figure B1 in appendix B for disposition details). Studies were excluded as off-topic or irrelevant (87); because they were literature reviews, book chapters, or summary articles rather than empirical studies (20); or because they focused on gender-based stereotype threat (107). The remaining 75 references were subject to a second round of screening to see whether they met the following criteria:

- Studied the effect of a social-psychological intervention (relevant to reducing the intensity of the psychological experience of stereotype threat) on improvements to student academic performance.

- Included Black students in the sample.
- Included K–12 students as the focus.

The second round of screening excluded 72 studies. Most studies (65) were excluded for failing to meet the first criterion—they explored various aspects of the negative impact of stereotype threat on performance rather than studying interventions to reduce the intensity of the experience of stereotype threat.

A second, broader verification search (using the broadest search term “stereotype threat” without the word “intervention”) was conducted to ensure that relevant studies had not been missed. No additional studies appropriate for inclusion were found among the 741 references identified.

*Assessing the quality of identified intervention studies.* The three remaining studies were subject to a final quality review to describe any methodological limitations, using a study coding protocol (see appendix C) based on the five criteria below from the *What Works Clearinghouse Procedures and Standards Handbook* (U.S. Department of Education 2008) for assessing the internal validity of studies examining the effects of interventions:

- *Outcome measures.* The measures used to assess impact must be shown to actually measure what they are intended to measure. The three studies reported on here used appropriate school measures of student achievement.
- *Random assignment process.* In experimental studies researchers use random assignment to assign

participants to experimental conditions (intervention or control) to ensure that the groups are as similar as possible on all characteristics so that the outcomes measured reflect the influence of the intervention only. Only one study had a limitation in this area (Good, Aronson, and Inzlicht 2003).

- *Attrition of participants.* Loss of participants can create differences in measured outcomes by changing the composition of the intervention or control groups. Both overall attrition and differential attrition (differences between intervention and control groups) are of concern. All three studies were acceptable in this area.
- *Intervention contamination.* Intervention contamination can happen when unintended events occur after intervention begins that could affect group outcomes and therefore the conclusions of the experiment. One study was noted as having a possible limitation in this area (Good, Aronson, and Inzlicht 2003).
- *Confounding factor.* It is important to examine factors beyond the intervention that might affect differences between groups, such as the effects of teachers or of the intervention provider more generally. No studies were noted as having problems in this area.

The completed study quality review protocols were used in developing the final list of limitations reported for each of the three studies.

For further details on the methodology see appendixes B and C.

that the real-world costs of stereotype threat are substantial. . . . If one could design effective interventions for reducing the experience of stereotype threat, then one would have a powerful tool for influencing an important set of societal problems.

The experimental manipulations used to study the effect of stereotype threat on academic test performance are of two kinds, direct and indirect. The direct way of inducing stereotype threat in experiments has been to tell the test-taking group that the test they will take has been sensitive to group differences in the past (for example, “this test shows racial differences”), thus raising the potential relevance of the stereotype as an explanation for poor performance. An indirect way of studying the negative effects of stereotype threat has been to inform students that a test is “diagnostic of your ability” (as in Steele and Aronson 1995), conveying that the test is designed to evaluate students’ performance along a stereotype-relevant trait (intellectual ability) and consequently bringing to the fore concerns about confirming the stereotype.

To the extent that stereotype threat might be a factor in some Black students experiencing extra stress when doing challenging academic work in school, what can be done to alleviate this stress and possibly improve their performance? Relatively few experimental studies have been conducted in classroom settings on interventions to explicitly reduce the experience of stereotype threat and thus im-

prove the academic performance of Black students. However, some recent classroom-based experimental studies were identified that have relevance for educators.

This study’s search for empirical studies of classroom-based social-psychological interventions designed to reduce stereotype threat and thus improve the academic performance of Black students initially identified 289 references (see box 1 and appendix B on

study methods). After relevant inclusion criteria were applied, three experimental studies were identified for description here. Those three studies are described in the following section.

### FINDINGS OF THREE EXPERIMENTAL STUDIES OF INTERVENTIONS TO REDUCE STEREOTYPE THREAT IN GRADE 7 CLASSROOM SETTINGS

All three studies reported on here found statistically significant positive effects of the tested interventions on achievement measures. The following intervention strategies were tested in the studies described in detail below:

- Reinforce for students the idea that intelligence is expandable and, like a muscle, grows stronger when worked.
- Teach students that their difficulties in school are often part of a normal “learning curve” or adjustment process, rather than something unique to them or their racial group.
- Help students reflect on other values in their lives beyond school that are sources of self-worth for them.

Table 3 at the end of the main report summarizes the outcome measures, analytic techniques, and the findings across the three studies. (Table B4 in appendix B summarizes the methodologies.)

Study 1: Blackwell, Trzesniewski, and Dweck 2007, “Implicit theories of intelligence predict achievement across an adolescent transition: a longitudinal study and an intervention”

#### *Intervention idea*

- Reinforce for students the idea that intelligence is expandable and, like a muscle, grows stronger when worked.

There is much research in psychology exploring the idea that some students can be trained to

**Intervention strategies in the studies described reinforced the idea that intelligence is expandable, taught students that difficulties in school are often part of a normal “learning curve,” or helped students reflect on other values in their lives as sources of self-worth**

think more productively about how they approach performance challenges. One belief that seems to affect how students approach such challenges is that intelligence is not fixed but malleable, that it can be developed through focus and effort and thus that intelligence can be taught (Dweck 1999; Whimbey 1975). Indeed, Aronson, Fried, and Good (2002) posit that some Black students might have developed a stereotype-consistent belief that their intellectual ability is “fixed,” causing them to feel more negative about academic performance situations than they would if they believed that their ability could grow with greater focus, effort, and creativity in problem-solving strategies. Alternatively, such students may feel that others see their ability as fixed and thus worry about negative inferences being drawn about them based on their performance. Thus, reinforcing the idea that intellectual ability is malleable and incrementally developed and that others view it in this way indirectly reduces students’ sense of psychological threat under challenging academic performance situations.

The first study reports on the effects of an intervention to teach students to see intelligence as incrementally developed rather than fixed.

*Research question.* Does teaching students to see intelligence as malleable or incrementally developed lead to higher motivation and performance relative to not being taught this theory of intelligence?

*Study sample.* The study sample included 91 grade 7 students in an urban public school with low-achieving students (52 percent Black, 45 percent Hispanic, and 3 percent White and Asian; 79 percent eligible for free or reduced-price lunch). There were 48 students in the intervention group and 43 in the control group. The two groups did not differ significantly in their prior academic achievement (fall term math grades) or on any baseline measures of motivation.

*What was the intervention?* Students in advisory classes—periods in the schedule when small

groups of students can receive more individual attention from a teacher—were randomly assigned to an intervention or control curriculum to test the effectiveness of teaching students about the theory of incremental intelligence. Both groups received eight weekly 25-minute sessions beginning in the spring of grade 7 during their regular advisory class period (to which they had been assigned at random by the school).

Both intervention and control groups received four 25-minute sessions on the brain, the pitfalls of stereotyping, and study skills. In four additional sessions the intervention group received information that focused on “growing your intelligence” and involved reading age-appropriate descriptions of neuroscience experiments documenting brain growth in response to learning new skills and class discussions on how learning makes students smarter. The intervention was based on previous experimental materials used in studies with college students (Aronson, Fried, and Good 2002; Chiu, Hong, and Dweck 1997). For these four sessions the control group received content unrelated to the malleability of intelligence and focused instead on topics about the brain and memory that were unrelated to the incremental theory of intelligence.

The sessions were delivered by 16 trained undergraduate assistants, with two undergraduates assigned to each class. To ensure consistent delivery of the intervention materials, session leaders received reading material and met weekly with the research team to review the material and prepare to present it to their assigned advisory class. Intervention and control workshop leaders met separately to train to prepare for the four sessions with different content.

*Results.* The researchers first provided results to show that their intervention had been successful

**The Blackwell, Trzesniewski, and Dweck study reports on the effects of an intervention to teach students to see intelligence as incrementally developed rather than fixed**

**The intervention group in the Blackwell, Trzesniewski, and Dweck study improved from pre- to postintervention, whereas the control group showed a continued downward trajectory in performance**

in teaching the intervention group students about the incremental theory of intelligence. The results from a theory of intelligence questionnaire given to students before and after the intervention showed that participants in the intervention group changed their opinions toward a more incremental view of intelligence after the intervention.

The researchers reported that a paired sample *t*-test (see box 2 for definition of key terms) was significant ( $t = 3.57, p < .05$ , Cohen's  $d = .66$ ), indicating that the intervention group endorsed the incremental theory more strongly after the intervention (mean score of 4.95 on the questionnaire) than before (4.36). The control group mean score on the questionnaire did not change (4.62 preintervention and 4.68 postintervention;  $t = 0.32$  and not significant, Cohen's  $d = .07$ ).

The important question, then, was whether achievement was higher in the intervention group as a result of the intervention. The researchers assessed the effect of the intervention on academic achievement by examining the growth curves of participants' math scores across three points in time: spring of grade 6 to fall of grade 7 (both prior to the intervention) and spring of grade 7 (postintervention). The researchers noted an overall downward trajectory in the mean math

scores for the entire sample (spring grade 6, 2.86; fall grade 7, 2.33; spring grade 7, 2.11). Analysis revealed a significant decline in scores for the total sample between the spring of grade 6 and fall of grade 7 ( $b = -.34, t = -4.29, p < .05$ ) and between fall of grade 7 and spring of grade 7 ( $b = -.20, t = -2.61, p < .05$ ).

The researchers further reported that the intervention group improved from pre- to postintervention (fall of grade 7 to spring of grade 7), whereas the control group showed a continued downward trajectory in performance (figure 1). That is, the intervention had a significant positive effect ( $b = .53, t = 2.93, p < .05$ ) on math scores from the fall of grade 7 to the spring of grade 7.

The researchers also collected comments from math teachers about students who had shown changes in motivational behavior after the advisory class sessions. (The teachers did not know to which condition their students had been assigned.) The study reported that 27 percent of the intervention group students received positive comments from math teachers about motivational change after the intervention, compared with 9 percent of the control group, a statistically significant difference.

*Methodological review.* No reservations were identified concerning the methodological quality of the study based on the study quality review protocol

#### BOX 2

##### Key terms

*t*-statistic. For a given sample size, the *t*-statistic indicates how often differences in means as large as or larger than those reported would be found when there is no true population difference in means (the "null hypothesis"). For example, a reported *t*-statistic that is statistically significant with a *p*-value of .05 indicates that in only 5 of 100 instances would this difference between the means in

a sample be found if the real population difference were zero.

*Degrees of freedom.* The number of independent observations used in a given statistical calculation and typically calculated by subtracting 1 from the number of independent observations (sample size).

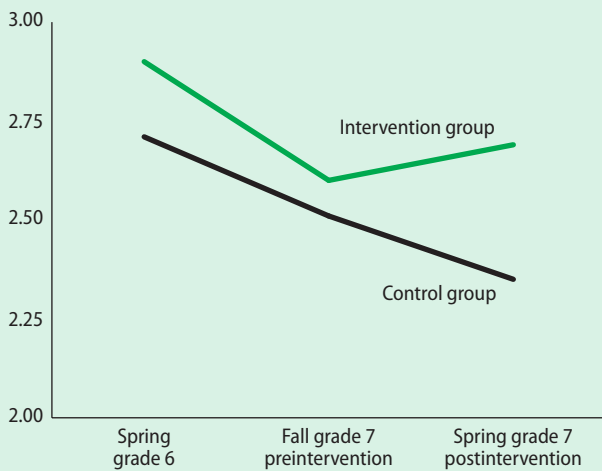
*b*-statistic. Represents the slope of a regression line based on predictors measured in their naturally occurring units.

*F*-statistic. Represents the ratio of the between-group variation divided by the within-group variation. A statistically significant *F*-statistic indicates that the mean is not the same for all groups (conditions).

*Effect size.* The impact of an effect expressed in standard deviation units.

*Cohen's d.* A type of effect size that represents the standardized mean difference between the treatment and control groups.

FIGURE 1

**Estimated mean math scores by experimental condition**

Source: Blackwell, Trzesniewski, and Dweck 2007.

criteria. (See summary of quality criteria on this study in appendix B.)

*Conclusions.* The researchers suggest that the incremental theory intervention “appears to have succeeded in halting the decline in mathematics achievement” (p. 258). Future research on the role of teachers in changing students’ beliefs about intelligence is needed, though these results are promising, particularly as the treatment was found to yield a significant effect in a low-income, urban setting where problems associated with minority underperformance can be severe.

*Study limitations.* This study was conducted in a single school, and thus the uniqueness of the school context or population as the setting for the intervention is unknown. Another limitation in generalizing the results of this study is that the sample of students was racially mixed (primarily Hispanic and Black), making it difficult to determine whether the intervention benefited both minority groups equally. The study authors acknowledge that the effects were measured at a single point in time, and it is not known whether the effects of the intervention would hold up for students as they moved to grade 8. The intervention sessions were delivered by trained undergraduate

assistants, not teachers. Thus, it is also unknown to what extent the intervention effect would hold up if delivered by teachers rather than trained undergraduates who, in this case, were closer in age to the students.

Study 2: Good, Aronson, and Inzlicht 2003, “Improving adolescents’ standardized test performance: an intervention to reduce the effects of stereotype threat”

### *Intervention idea*

- Teach students that their difficulties in school are often part of a normal “learning curve” or adjustment process, rather than something unique to them or their racial group.

A related potentially unproductive thought process occurs when students attribute academic struggles to their intellectual limitations, which may be more likely for students who struggle with stereotypes about their group’s intellectual inferiority. To the extent that students attribute normal difficulties—for instance, those that occur with hard-to-learn topics or concepts—to fixed personal inadequacies, they may experience more distraction, anxiety, and pessimism. Thus, interventions might reduce the negative effects of stereotype threat, as well as other forms of doubt, by encouraging students to attribute difficulty in school to the transitory struggles all students experience.

*Research question.* Can teaching students to attribute academic difficulties to transitory situational causes rather than to stable personal causes improve standardized math and reading test scores?

*Study sample.* The study took place in a rural school district in Texas serving a largely low-income and predominantly minority

**The Good, Aronson, and Inzlicht study asks whether interventions might reduce the negative effects of stereotype threat, as well as other forms of doubt, by encouraging students to attribute difficulty in school to the transitory struggles all students experience rather than to fixed personal inadequacies**

population (63 percent Hispanic, 15 percent Black, and 22 percent White). Study participants were 138 grade 7 students enrolled in a computer skills class as part of their junior high school curriculum. 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 regular course curriculum, students learned a variety of computer skills including using email and designing web pages.

*What was the intervention?* Shortly after the school year began (mid-October), students in the computer skills class were randomly assigned a mentor with whom they communicated in person and by email throughout the school year. They were also randomly assigned to receive one of four types of educational messages from their mentors:

- *Incremental message (40 students).* Students learned about the expandable nature of intelligence (as explored in the previously described study).
- *Attribution message (36 students).* Students learned about the tendency for all students to initially experience difficulty during grade 7 and about the tendency for this difficulty to subside with time and for performance to improve.
- *Combination of messages (30 students).* Students received both the incremental and attribution messages.
- *Control condition (32 students).* Students learned about the perils of drug use (an unrelated topic).

The mentors conveyed the content of their assigned messages in person to the students during two school visits of 90 minutes each. After learning this information, students created public service announcements on the web with guidance from their mentor,

reinforcing the message that they had learned and helping to internalize the message through a self-persuasion process. A restricted web space was created for each of the four conditions so that students learning a particular message could read more about their assigned message but not read the messages for the other three groups and acquire additional ideas for polishing their web page.

The mentors were 25 college students who participated in a three-hour training session on mentoring required by the district and then supplementary training by the researchers on how to convey the four messages tied to the four conditions in the study. The same mentors delivered the intervention to students in three of the four conditions.

*Results.* At the end of the school year participating students' scores on statewide standardized tests in math and reading were analyzed for the four groups of students.

Math test scores were analyzed using a 2 (gender) by 4 (experimental condition) analysis of variance. The math analyses are not presented here because they focused on understanding gender effects, which were not the focus of this report.

Reading scores on the Texas Assessment of Academic Skills were analyzed using a one-way analysis of variance that compared the performance of students participating in the four experimental conditions. Although the researchers were interested in differences between Black and White students' performance in the four conditions, the samples were not large enough to analyze the two groups separately. The analysis of variance conducted on the state reading test scores revealed a significant effect ( $p < .05$ ) of the conditions,  $F(3,125) = 2.71$ . Follow-up statistical tests showed that scores on the state reading test were significantly higher for students in the conditions receiving the incremental (malleable) intelligence message (mean score of 88.26) and the attributional message (89.62) than for students in the control group (84.38) (table 1). There was

**Statistical tests for the Good, Aronson, and Inzlicht study showed that scores on the state reading test were significantly higher for students in the conditions receiving the incremental (malleable) intelligence message and the attributional message than for students in the control group**

TABLE 1

**Reported intervention impacts on the spring grade 7 state reading test (Texas Assessment of Academic Skills)**

Intervention effect	Incremental condition	Attribution condition	Combined condition	Control condition
Mean reading score	88.26	89.62	86.71	84.38
Standard deviation	7.17	7.01	8.70	7.79
Difference between each experimental condition and control condition	$t(65) = 2.07$ $p < .041$ Cohen's $d = .52$	$t(61) = 2.72$ $p < .008$ Cohen's $d = .71$		Not significant

Source: Good, Aronson, and Inzlicht 2003.

no significant difference between the combined messages condition and the control condition.

*Methodological review.* Applying the study quality review criteria revealed two limitations of the methodology (see appendix B for complete summary).

*Random assignment process.* The study reported that 6 of the 138 students' scores were removed from the analysis, which could be considered a disruption in the random assignment process. In addition, no evidence was presented of the equivalence of the four groups on baseline achievement. Although the authors reported that these six students did not come from any particular experimental condition or group, it is difficult to know how well the random assignment process worked in creating equivalent groups at baseline without these data. Therefore, the study results showing differences between experimental conditions after the treatment should be interpreted with caution.

*Intervention contamination.* The same mentors delivered the intervention to students in three of the four conditions, so the intervention conditions could have been somewhat blurred if the mentors brought knowledge from one condition to their delivery of another. However, under What Works Clearinghouse review standards, contamination such as occurred in this study is not considered grounds for downgrading a study.

*Conclusions.* The authors suggest that showing the positive impact of these attitude-changing

interventions on state test performance builds on prior experimental studies showing the effects of similar interventions on college students' classroom performance (see Wilson and Linville 1985).

*Study limitations.* The sample in this study was mixed. Although it consisted mainly of minority students, Hispanic students made up 63 percent of the sample and Black students only 15 percent. So, there are limitations in generalizing the findings to Black students alone. As in the first study, teachers did not deliver the intervention and thus it is difficult to know under what conditions teachers can effectively deliver the intervention (for instance, how much teacher training would be needed, what kind of materials would they use). Nevertheless, the results are interesting, especially the finding of the intervention conditions' significant effect on academic achievement in a low-income school setting.

---

Study 3: Cohen, Garcia, Apfel, and Master 2006, "Reducing the racial achievement gap: a social-psychological intervention"

#### *Intervention idea*

- Help students reflect on other values in their lives beyond school that are sources of self-worth for them.

Another route to alleviating stereotype threat is to allow individuals to affirm an alternative positive identity—one that shores up their sense of self-worth in the face of threat. Through

**The Cohen et al. study examined whether allowing students to affirm an alternative positive identity one that shores up their sense of self-worth in the face of threat would alleviate stereotype threat**

self-affirmation people reinforce their sense of personal worth or integrity by reflecting on sources of value and meaning in their lives (Steele 1988). People are better able to tolerate psychological threat in one domain (such as school) if they can shore up their self-worth in another domain (such as family). Laboratory research shows that self-affirmations can reduce stress (Creswell et al. 2005).

For example, college students asked to give a speech in front of a sullen audience displayed lower levels of the stress hormone cortisol if they were first given the opportunity to engage in the self-affirmation exercise of reflecting on an important value, such as their relationships with friends.

*Research question.* Would Black students perform significantly better in a targeted course when they received a self-affirmation intervention than when they did not? The researchers hypothesized less of an intervention effect for the nonstereotyped group, as the risk factors (elevated stress and psychological threat) were expected to be lower for nonstereotyped students, who do not contend with a negative stereotype about their racial group.

*Study sample.* The researchers report the results of two randomized experiments. The second, a replication study, took place a year after the first study and with a different cohort of students. A total of 119 Black and 124 White grade 7 students participated in the two studies (roughly evenly distributed across the two studies). Students were from a suburban northeastern middle school. The three teachers who participated all taught the same subject area. At the beginning of the fall semester students were randomly assigned to an intervention or control condition. Teachers were unaware of which students in their classes were assigned to which of the two conditions, and the two experimental conditions as described below were presented to students as part of the regular classroom curriculum.

*What was the intervention?* The intervention was intended to engage students in a self-affirmation process that would alleviate some of the stress Black students might feel from stereotype threat and thereby improve their academic performance. The affirmation intervention was a series of writing assignments designed to induce feelings of self-worth and test whether psychological threat could be lessened through asking students to “reaffirm” their “self-integrity.” The assignments (developed by the researchers) were provided to students in an envelope and included self-explanatory instructions that required little teacher involvement. The teachers’ role in the study was to hand out the envelopes containing the writing assignments, provide a brief scripted introduction to students, and then to remain at their desks and allow students to independently complete the assignment and return their work to the teacher in the envelope.

The envelopes were identical for the intervention condition and the control condition assignments, so teachers were unaware of which students were receiving the self-affirmation intervention. The self-affirmation assignment was designed to encourage students to think about a personal value or values they had singled out as important and its significance in their lives.

Students in both groups received a list of values and were asked to read and think about them. The values were notions such as athletic ability, creativity, music, relationships with friends, independence, religious values, and sense of humor. The instructions for students in the intervention group asked them to select their most important value (or values) and to write a paragraph about its importance to them. The instructions for students in the control group asked them to select their least important value (or values) from the list and write about why it might be important to someone else. The instructions then asked the students in the intervention group to write the top two reasons why the value (or values) they selected was important to them. The students in the control group were instructed to write the top two reasons why someone else might consider their least important value important. Finally, the



instructions asked students to select their level of agreement with four statements about the values they chose (most important value for the intervention condition and least important for control condition) as a way of reinforcing their value selection in the affirmation condition.

Teachers presented the instructions to students as a regular classroom assignment. Completing the assignment took students in both intervention and control conditions about 15 minutes. One structured writing assignment was provided to students in the first study, and two were provided to students in the replication study.

*Results.* The outcome data collected were students' GPAs from official transcripts in the targeted course for the fall term in which the intervention was delivered. The data were analyzed using multiple regression. The interaction of race (Black or White) and experimental condition (affirmation intervention condition or control condition) was significant for study 1 ( $b = 0.29, t(98) = 2.00, p < .05$ ) and study 2 ( $b = 0.52, t(119) = 2.80, p < .01$ ), as was the treatment main effect for Black students in study 1 ( $b = 0.26, t(41) = 2.44, p < .02$ ) and study 2 ( $b = 0.34, t(60) = 2.69, p < .01$ ). Black students receiving the affirmation intervention had higher grades in the targeted course in the fall term than did Black students in the control condition. The difference in GPA for Black students in the intervention condition and the control condition was 0.26 point in the first study and 0.34 point in the second replication study.

The mean differences in the outcome measure for Black students and White students by three levels of prior academic performance are shown in table 2. The study reports that the intervention was as strong for previously low-performing Black students ( $t(31) = 2.74, p < .01$ ) as for previously moderate-performing Black students ( $t(30) = 2.40, p < .02$ ). The previously high-performing Black students benefited less from the intervention condition ( $t(31) = 1.72, p < .10$ ).

The intervention effect on the difference in GPA between Black students receiving the affirmation intervention and those in the control group was 0.43 point for the previously low-performing group, 0.44 point for the previously moderate-performing group, and 0.22 point for the previously high-performing group. In all three cases Black students who received the affirmation intervention had a higher mean GPA in the course than did Black students in the control group. Additionally, the intervention effect for Black students extended to courses beyond the targeted course, as evidenced in an analysis of students' mean GPA in core academic courses.

Combining data from studies 1 and 2 shows that the intervention reduced the percentage of Black

**In the Cohen et al. study Black students receiving the affirmation intervention had higher grades in the targeted course in the fall term than did Black students in the control condition**

TABLE 2

**Covariate-adjusted mean grade point average (averaged over both studies) for intervention and control groups, by level of preintervention performance**

Condition	Low performing student group		Moderate performing student group		High performing student group	
	Black	White	Black	White	Black	White
Affirmation intervention	1.7	2.2	2.8	3.3	3.5	4.0
Control	1.3	2.3	2.4	3.2	3.3	4.0

*Note:* The grade point average is that received in the fall term in the academic subject in which the experiment was carried out at the beginning of the school year. The academic subject area was not identified in the study except to say that it was not one that was typically related to gender stereotypes (for example, math).

Source: Cohen et al. 2006.

students earning a D or below in the fall term of the course from 20 percent in the control group, a rate consistent with historical norms at the school, to 9 percent in the intervention group, a significant difference (figure 2). There was no significant difference between the intervention and the control conditions for White students.

*Methodological review.* No reservations were identified concerning the methodological quality of this study on the study quality review protocol criteria (see appendix B for details).

*Conclusions.* The authors conclude that “our intervention is among the first aimed purely at altering psychological experience to reduce the racial achievement gap.” That is, rather than “lift all ships,” the intervention benefits those most in need—low-performing Black students. Additionally, “the research highlights the importance of situational threats linked to group identity in understanding intellectual achievement in real-world, chronically evaluative settings . . . [and] challenge[s] conventional and scientific wisdom by demonstrating that a psychological intervention,

although brief, can help reduce what many view as an intractable disparity in real-world academic outcomes” (p. 6).

*Study limitations.* Limitations of the study include the fact that it was conducted in only one school and grade level in a suburban district and that it is difficult to determine how representative the sample is of the general population from which it was drawn. It is thus difficult to know whether the intervention would yield similar benefits in other schools of varying demographic and socioeconomic characteristics and in other grade levels. Additionally, as with the other two interventions reported here, it is unclear whether the intervention would be similarly beneficial when prepared and implemented entirely by teachers rather than trained researchers. Still, the results are promising, as the intervention effect proved replicable (obtained in two separate studies), and the effect of the intervention on minorities’ grades was consistently positive across most of the range of prior achievement.

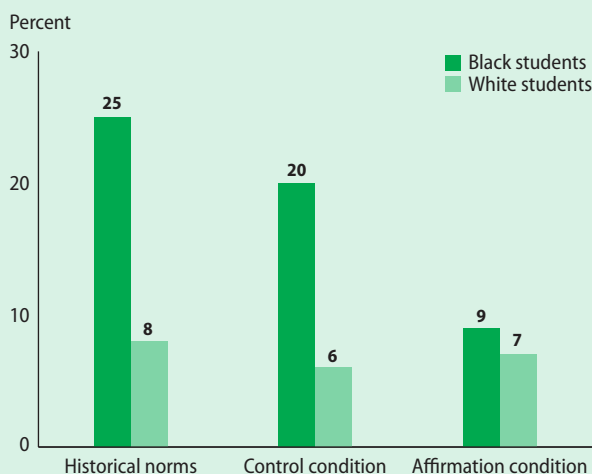
## CONCLUDING THOUGHTS ON TURNING RESEARCH INTO PRACTICE

The objective of this report was to conduct a systematic search to identify classroom-based strategies designed to reduce stereotype threat and thus to improve the academic performance of Black students. The three studies that were identified found that the following social-psychological strategies had impacts on minority group achievement:

- Reinforce for students the idea that intelligence is expandable and, like a muscle, grows stronger when worked.
- Teach students that their difficulties in school are often part of a normal “learning curve” or adjustment process, rather than something unique to them or their racial group.
- Help students reflect on other values in their lives beyond school that are sources of self-worth for them.

FIGURE 2

**Percentage of Black and White students receiving a grade of D or lower in targeted course in same semester as the intervention, by experimental condition**



Source: Cohen et al. 2006.

When considering these studies, several limitations of this review are important. First, the search was very focused, intended to identify only studies of interventions that had been tried in real school settings. For each strategy, there is a larger body of social-psychological theory and research that led to the testing of the particular intervention that is not reviewed. Few social-psychological studies are conducted in classroom settings, but it was important to focus only on studies with possible applicability for educators. Another limitation is that these strategies do not represent all the possible ways of reducing stereotype threat, only those that have been studied with rigorous research. There may be other, better ways of reducing stereotype threat that have not been studied.

Finally, readers should be aware that the studies here are small in scope, and their replicability is unknown. However, it is clear that the stereotype threat phenomenon has been experimentally shown to exist across a wide variety of studies. Thus, it is important to share ideas for reducing the negative effects of this phenomenon, even if they are in the early stages of knowledge development. For the three experiments reported on here, evidence suggests that such strategies might reduce the level of psychological threat some Black students feel in the classroom and that, combined with other efforts, these strategies could benefit the performance of Black students.

Although researchers have developed specific protocols to follow for the interventions in some contexts, educators might need to adapt the interventions to fit their classrooms and then monitor them to determine what impact they have. An understanding of the purpose and process involved in using the strategy is important, as is professional wisdom about how to apply the process in a given classroom context. Such understanding and awareness help ensure that the spirit of the intervention is not lost when local conditions prevent a teacher from strictly following the protocols. If school teams or teachers do not grapple with the underlying rationale or purpose

of an intervention, key elements may be left out, rendering the intervention less effective.

For example, the timing of interventions is important. The interventions in the Blackwell, Trzesniewski, and Dweck (2007) and Cohen, Garcia, Apfel, and Master (2006) studies seemed to halt or at least slow a downward performance spiral for students. All three studies were conducted on students in grade 7, which raises the possibility that there may be windows of opportunity for influencing student attitudes and beliefs. For instance, grade 7 is a time when concerns about race-based stereotype increase for minority students and is a developmental period when adolescents' sense of identity is in flux. Interventions may be particularly influential at such junctures by altering students' early trajectory and preventing a path of compounding failures.

Thus, the grade level at which the intervention ideas are applied is an important consideration, as is the timing during the year. For example, the self-affirmation assignment may be most effective when given at times of high stress, such as the beginning of the school year, to halt or reverse a downward slide that could otherwise feed off itself, with stress worsening performance and with deteriorating performance heightening stress in a repeating cycle. Such downward slides coincide with academic transitions, such as the transition to middle school, high school, or college. These are times when performance standards shift upward, when students' sense of identity is not yet crystallized, and when social-support circles are disrupted, heightening stress and feelings of exclusion. If a small psychological intervention can interrupt a downward spiral at such times, or prevent it from emerging, there is the possibility of large and long-term effects (Cohen et al. 2006).

**The evidence suggests that strategies such as those analyzed in the three experiments reported on here might reduce the level of psychological threat some Black students feel in the classroom and that, combined with other efforts, these strategies could benefit the performance of Black students**

**The three studies reported here suggest that seemingly small actions in the classroom when well timed, well targeted, and thoughtfully and systematically implemented can produce positive results for minority students**

Social-psychological research suggests that human intellectual performance and motivation are fragile (Aronson and Inzlicht 2004; Aronson and Steele 2005). The three studies reported here suggest that seemingly small actions in the classroom—when well timed, well targeted, and thoughtfully and systematically implemented—can produce positive results for minority students.

It is important to bear in mind, however, that none of these interventions would work unless students already have some ability or motivation to improve academically and unless the school has the foundational resources to permit students to achieve at a higher level. The interventions will

not teach a student to spell who does not already know the fundamentals. They will not suddenly motivate an unmotivated student or turn a low-performing and underfunded school into a model school. More generally, the interventions would not work if there were not broader positive forces in the school environment (committed staff, quality curriculum) operating to facilitate student learning and performance. Without these broader positive forces, social-psychological interventions, while potentially reducing psychological threat levels for some students, would be unlikely to boost student learning and achievement. However, when these broader positive forces are in place, social-psychological interventions such as those reported on here may help Black and other minority students to overcome stereotype threat and improve their performance in school.

TABLE 3

**Summary of effects reported by the three studies**

Study	Outcome measure	Analysis technique	Treatment effect <sup>a</sup>	Description of difference between intervention and control groups
Blackwell, Trzesniewski, and Dweck (2007)	Predicted math grades on a 0 [F] to 4.0 [A] scale	Growth curve analysis	$t(371) = 2.93, p < .05$	According to a figure presented in the study report, the intervention group averaged roughly a 0.10 increase in math course grades from fall to spring. The increase does not represent a letter grade change (such as C+ to B-); it remains within the C+ range. However, the increase for the intervention group from fall to spring, though small, contrasted with a decline in grades for the control group from fall to spring (from roughly a C+ to a C).
Good, Aronson, and Inzlicht (2003)	Reading achievement scores on the Texas Assessment of Academic Skills (TAAS) standardized tests	Analysis of variance	Incremental condition: $t(65) = 2.07, p < .05$ Attributional condition: $t(61) = 2.72, p < .01$ Effect sizes reported: <ul style="list-style-type: none"> <li>Between students who received the message on the incremental theory of intelligence and students in the control group: Cohen's <math>d = .52</math></li> <li>Between students who received the attributional message and students in the control group: Cohen's <math>d = .71</math></li> </ul>	Compared with students in the control condition, students in the incremental condition earned an average 3.88 points higher on the TAAS and students in the attributional intervention condition an average of 5.24 points higher.  The .52 and .71 effect sizes reported are considered moderate to large effects for educational interventions.
Cohen, Garcia, Apfel, and Master (2006)	Targeted course grade point average, on a 0 [F] to 4.33 [A+] grade point average scale	Multiple regression	Study 1: $t(41) = 2.44, p < .02$ Study 2 (replication): $t(60) = 2.69, p < .01$	According to the authors, the intervention effects translated into an estimated 0.26 point increase in study 1 and 0.34 point increase in study 2, respectively, in fall targeted course grades for Black students in the intervention condition compared with those in the control condition.

*Note:* Only the Cohen et al. (2006) study directly analyzed the reduction in the achievement gap between Black and White students. The other two studies reported on positive effects of the intervention on the overall sample of students, which included primarily minority (Black and Hispanic) students. However, the two studies were not able to compare minority student improvement with that of White students.

a. The effect size statistic represents the impact of the effect in standard deviation units. Because only the Good, Aronson, and Inzlicht (2003) study calculated and reported effect sizes, effect sizes could not be compared across the studies. Instead  $t$ -statistics and corresponding  $p$ -values are reported. For a given sample size, the  $t$ -statistic indicates how often differences in means as large as or larger than those reported would be found when there is no true population difference in means (the null hypothesis). The number in parentheses with the  $t$ -statistic indicates the degrees of freedom. Cohen's  $d$ , a type of effect size, represents the standardized mean difference between the intervention and control groups. It is calculated by dividing the difference between the intervention group and control group means by either their average standard deviation or by the standard deviation of the control group. See box 2 for more detailed definitions.

*Source:* Authors' compilation and calculation from Blackwell, Trzesniewski, and Dweck (2007); Good, Aronson, and Inzlicht (2003); and Cohen et al. (2006).

---

**NOTES**

1. This report uses the term *Black students* throughout, even when the reported study used a different term.
2. The No Child Left Behind Act of 2001 refers to “scientifically based research” as an important criterion for educators as they consider new interventions or strategies. Randomized controlled experiments are said to be the “gold standard” of the sciences, the highest standard of evidence or methodology available for studying the effectiveness or impact of an intervention. In such experiments participants are randomly assigned to one of two or more conditions that differ in a critical way that is hypothesized to have a particular impact. At the simplest level there is an

intervention group that receives the intervention and a control group that does not. If the students randomly assigned to the intervention group perform significantly better on the outcome measure than do students in the control group (a less than 5 percent probability of the difference between the two groups being due to chance), it is likely that the difference in performance was the result of the intervention. Random assignment creates groups that should be (on average) identical in all dimensions except for receiving the intervention; thus, any differences in outcomes can be attributed to the intervention. The three published studies identified and examined in this report use this type of research design for testing interventions to reduce stereotype threat in classrooms and improve academic performance.

---

## APPENDIX A RESEARCH ON THE RELATIONSHIP BETWEEN STEREOTYPE THREAT AND BLACK STUDENTS' ACADEMIC PERFORMANCE

Students' academic performance in classrooms, because of processes such as stereotype threat, can be more variable than people customarily think, fluctuating with changes in the situation (Aronson and Steele 2005). For example, studies show that women's performance on math tests can be made to rise and fall with surprising ease. When women were asked to generate a short list of qualities shared by men and women, their math test performance rose (Rosenthal and Crisp 2006). In another study, when women were reminded that they were students at a selective liberal arts college and their attention was thus turned away from their gender, women's spatial-abilities test performance rose and the male-female gap shrank (McGlone and Aronson 2006). When women took a test in the presence of men, their math performance declined (Inzlicht and Ben-Zeev 2000). But when women were presented with a female test proctor who excelled in math, their performance improved and the male-female gap again shrank (Marx and Roman 2002). Such studies underscore the degree to which human performance is shaped by environmental and psychological forces—not simply by how smart a student is or how hard he or she works.

Research on stereotype threat began with laboratory studies exploring why Black college students seemed to be performing below their potential. Although a test-taking situation may seem objectively the same for all students, some students, because of their social identity, may experience it in a very different way.

Steele and Aronson (1995) conducted an experiment to explore the negative impact of administering a test under potentially stereotype-threat-inducing conditions by randomly assigning study participants to two different test-taking conditions. In one test-taking condition a standardized test (composed of verbal Graduate Record Exam

items) was presented to one group of college students as “diagnostic of intellectual ability.” It was hypothesized that Black students in this condition would worry that performing poorly could confirm a stereotype about their racial group's intellectual ability. Black students performed worse in this condition than when the same test was given in a second condition that introduced the test as one that was “not diagnostic of your ability.” The two ways of introducing the test had no effect on the performance of White students. Black students in the study sample answered roughly 8 of 30 test items correctly in the “threat” condition and roughly 12 of 30 correctly in the “no threat” condition.

Since this first Steele and Aronson study, the concept of heightened performance stress or anxiety for certain groups has been found across a variety of potential stereotypes and minority groups. Experimental studies have shown that detrimental stereotype threat affects not only Black students on verbal tests, but Hispanic students on verbal tests (Aronson 2002), young women on math tests (Quinn and Spencer 2001; Spencer, Steele, and Quinn 1999), White men in certain sports situations (Stone et al. 1999), students from socioeconomically disadvantaged households on school tests (Croizet and Claire 1998), and high-performing White students on math tests when they are reminded of the stereotype of Asian superiority in math (Aronson et al. 1999).

---

### Direct and indirect manipulations of stereotype threat

Experimental manipulations of stereotype threat have differed, and these differences can be relevant to test-taking instructions used in K–12 settings (Quinn and Spencer 2001). One direct way of inducing stereotype threat in experiments has been to tell the test-taking group that the test they will take has been sensitive to group differences in the past (for example, “this test shows racial differences”), thus raising the potential relevance of the stereotype as an explanation for the test taker's poor performance. Although drawing attention to group differences just before administering a test

(for example, stating that girls have performed worse than boys on the math test in the past, or that Black students as a group performed poorly on the test the previous year) could cause a few students to rise to the challenge, the laboratory research suggests that the average performance of negatively stereotyped group members decreases. The fact that some of this laboratory research was conducted with college students on elite campuses (Steele and Aronson 1995) suggests that such a detrimental effect could occur even among the most confident and skilled students.

A less direct way of studying the negative effects of stereotype threat has been to inform the students in the study that the test is “diagnostic of your ability” (as in Steele and Aronson 1995). This conveys that the test is designed to evaluate students’ performance along a stereotype-relevant trait (intellectual ability) and consequently can bring to the fore concerns about confirming the stereotype. Experimental studies have shown that the performance of the stereotyped group tended to be poorer in the group that received the instruction that the test was diagnostic of ability than in the comparison group that received instructions emphasizing that the test is not diagnostic of ability (Spencer et al. 1999; Steele and Aronson 1995).

The power of these direct and indirect ways of inducing stereotype threat relates to a general psychological principle that has been widely studied—the priming effect. The priming effect refers to the tendency for people to conform their thoughts, feelings, and behaviors to psychologically accessible mental constructs such as stereotypes. Thus, when individuals are “primed” with a negative stereotype, their interpretations of ambiguous stimuli, behaviors, or performances are often influenced by the stereotype, even when the priming occurs at the unconscious or subliminal level. The implication of priming effects for teachers trying to encourage their students to perform to their potential is that subtle events in the classroom can undermine a student’s confidence, trust, and performance. Studies also show that priming positive concepts, such as being a

good student, can improve performance (McGlone and Aronson 2006).

---

### Mediating mechanisms

Although inducing stereotype threat conditions has been shown across multiple studies to result in poorer performance from the stereotyped group, the research has been less clear on the mediating mechanisms—on why stereotype threat results in poorer performance.

Some researchers have studied mediating mechanisms that might interfere with the quality of the performance under conditions of stereotype threat such as increases in stress, anxiety, self-consciousness, mental load, or heightened demands on working memory—all of which could lead to less focus on the task at hand, suboptimal test-taking strategies (such as guessing more), and underperformance (Beilock et al. 2006; Schmader and Johns 2003). Making students aware of the effects of anxiety from stereotype threat has been shown in several studies to improve the performance of negatively stereotyped students (Johns, Schmader, and Martens 2005; McGlone and Aronson 2007), presumably because awareness of external pressures reduces the tendency to attribute test anxiety to one’s intellectual shortcomings by providing an alternative attribution. The study findings suggest that helping students understand stereotype threat might inoculate them in some way against the extra stress or lack of focus that might take their attention away from the performance at hand.

---

### Experiencing stereotype threat over time

Although difficult to study, some long-term effects of repetitively experiencing the extra stress due to stereotype threat have been suggested. One consequence might be that as Black students have the opportunity to make choices in school, some of them might avoid challenges by selecting easier courses or assignments when they are being academically evaluated. Studies with middle school minority students have found that students asked



for easier problems to solve when confronted with the prospect of being intellectually evaluated on the basis of their performance (Aronson and Good 2002). Compared with White students, the minority students showed a strong tendency to take on less challenging work, presumably because they were threatened by the prospect of looking less intelligent if the challenge proved too great.

But there were individual differences that moderated these findings. Minority students were less likely to avoid a challenge if they believed that the challenge could increase their intelligence. Additionally, reducing stereotype threat through an experimental intervention increased minorities' interest in taking challenging rather than easy college courses (Walton and Cohen 2007).

**APPENDIX B  
METHODOLOGY**

The methodology for this study included a systematic search, screening, and review process to ensure methodological replicability.

**Search process**

A systematic search was conducted to identify empirical studies of classroom-based social-psychological interventions designed to reduce stereotype threat and thus to improve the academic performance of Black students.

The broadest search used the Education Resources Information Center (ERIC) and the search term “stereotype threat,” resulting in 44 citations. Subsequently, narrower search term combinations, such as “stereotype threat” and “intervention,” and “achievement gap” and “intervention,” were used to search several bibliographic databases. To identify new literature, PsycInfo was used to search on “stereotype threat” and “social identity

threat.” Forward citation searches using seminal stereotype threat papers and searches of reference lists in newly published work were also conducted. The searches yielded 158 citations (table B1). In addition, a web site on this topic, with an extensive reference list of peer-reviewed journal articles, was reviewed ([www.reducingstereotypethreat.org](http://www.reducingstereotypethreat.org)). Launched on November 28, 2007, the web site was developed by Steve Stroessner (Columbia University) and Catherine Good (Baruch College), but is now maintained solely by Stroessner. Until June 26, 2008, it was updated monthly or bimonthly. Scanning the web site reference list resulted in an additional 131 citations, for a total of 289 references.

**Screening**

The references were screened twice, first for content relevance and then for intervention and sample relevance (see appendix C for the six screening criteria).

*Initial screening of references.* Citation information from these 289 references was entered into an

TABLE B1  
**Search results**

Search engine or web site	Database	Search terms	Number of references identified
ERIC		stereotype threat	44
EBSCOhost	PsycINFO	achievement gap and intervention	0
EBSCOhost	Academic Search Premier	achievement gap and intervention	0
Wilson Web	Education Index	racial achievement gap and intervention	0
EBSCOhost	PsycINFO	stereotype threat and intervention	3
EBSCOhost	Academic Search Premier	stereotype threat and intervention	1
Wilson Web	Education Index	stereotype threat and intervention	0
EBSCOhost	ERIC	stereotype threat and intervention	2
First Search	Dissertation abstracts	stereotype and threat	108
First Search	Dissertation abstracts	stereotype threat and intervention	0
First Search	Dissertation abstracts	racial achievement gap and intervention	0
<a href="http://www.reducingstereotypethreat.com">www.reducingstereotypethreat.com</a>	na	na	131
<b>Total references</b>			<b>289</b>

na is not applicable.

Source: Authors' compilation.

internal tracking database for documenting disposition. These references were first screened for inclusion using three questions on content relevance (see article screening protocol in appendix C):

- Is the article on topic?
- Is the citation an empirical study?
- Does the study focus on race-based stereotype threat?

If the title or abstract did not provide enough information about the study, the full article was reviewed for relevance. Table B2 and figure B1 show the disposition of references.

Applying the first set of three criteria in the article screening protocol led to 214 exclusions:

- 87 references, as off-topic or irrelevant.
- 20 references, which were literature reviews, book chapters, or summary articles—not empirical studies.
- 107 references, which focused on gender-based stereotype threat (conditions under which women perform worse than men on math tests) rather than race-based stereotype threat.

*Second-level screening of relevant references.* The remaining 75 references were subject to a second round of screening to determine whether the studies met the following criteria:

- Examined the effect of a social-psychological intervention (relevant to reducing the intensity of the psychological experience of stereotype threat) on improvements to student academic performance.
- Included Black students in the sample.
- Included K–12 students as the focus (not college students fulfilling requirements to participate in experiments).

FIGURE B1

### First- and second-level screening and assessment of the quality of studies

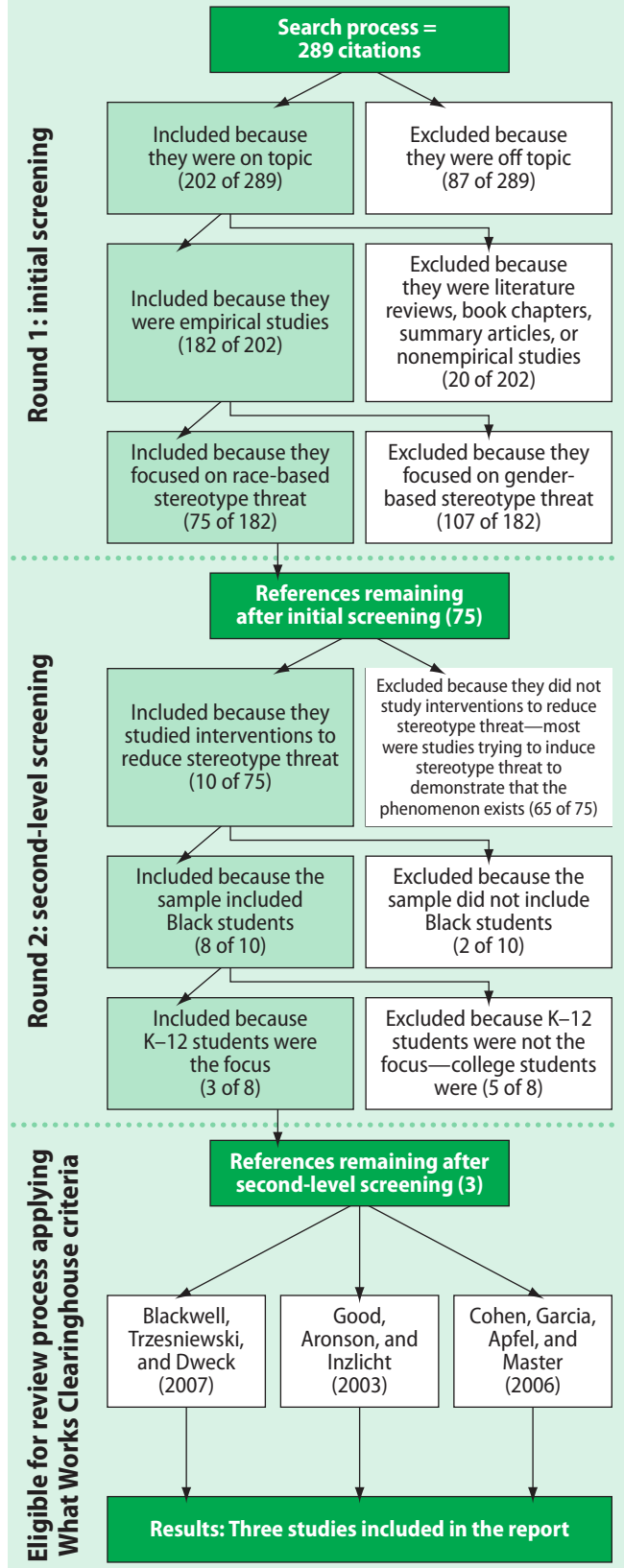


TABLE B2  
Disposition of references

Search engine	Database	Search terms	Number of studies identified	Round 1			Round 2			Total references included after screening
				Off topic	Nonempirical (chapter/literature review)	Gender stereotype threat only	Not an intervention to reduce stereotype threat	Sample excludes Black students	Non K 12 classroom setting/student focus	
ERIC		stereotype threat	44	10	10	18	5	0	0	1
EBSCOhost	PsycINFO	achievement gap and intervention	0	0	0	0	0	0	0	0
EBSCOhost	Academic Search Premier	achievement gap and intervention	0	0	0	0	0	0	0	0
Wilson Web	Education Index	racial achievement gap and intervention	0	0	0	0	0	0	0	0
EBSCOhost	PsycINFO	stereotype threat and intervention	3	0	0	1	0	0	0	2
EBSCOhost	Academic Search Premier	stereotype threat and intervention	1	0	0	1	0	0	0	0
Wilson Web	Education Index	stereotype threat and intervention	0	0	0	0	0	0	0	0
EBSCOhost	ERIC	stereotype threat and intervention	2	1	1	0	0	0	0	0
First Search	Dissertation abstracts	stereotype and threat	108	35	2	32	36	1	2	0
First Search	Dissertation abstracts	stereotype threat and intervention	0	0	0	0	0	0	0	0

(CONTINUED)

TABLE B2 (CONTINUED)  
Disposition of references

Search engine	Database	Search terms	Number of studies identified	Round 1			Round 2			Total references included after screening
				Off topic	Nonempirical (chapter/literature review)	Gender stereotype threat only	Not an intervention to reduce stereotype threat	Sample excludes Black students	Non K 12 classroom setting/student focus	
First Search	Dissertation abstracts	racial achievement gap and intervention	0	0	0	0	0	0	0	
www.reducingstereotypethreat.org	na	na	131	41	7	55	24	1	3	0
Total references excluded	na	na	na	87	20	107	65	2	5	na

na is not applicable.

Source: Authors' compilation.

This second round of screening excluded 72 studies (see table B2).

The majority of the studies (65) were excluded for not meeting the first criterion. The studies explored various aspects of the negative impact of stereotype threat on Black students. They did not test a social-psychological intervention aimed at improving Black student performance by reducing stereotype threat or mitigating its effects.

Two studies were excluded because they did not include Black students. Studies that included Black students as part of their sample were retained. No specific percentage of the sample was stipulated as having to be Black students. (Also, no criterion was specified for sufficient representation of Black students for analyses of outcomes by ethnicity.)

Of the three studies that remained after screening, only one study (Cohen et al. 2006) specifically analyzed race as a factor. In the Blackwell, Trzesniewski, and Dweck (2007) study the students were from a large urban school district, and all were minority (52 percent were Black and 45 percent were Hispanic). In the Good, Aronson, and Inzlicht (2003) study, the students were from a rural district in Texas with 70 percent eligible for free or reduced-price lunch (67 percent were Hispanic, 13 percent were Black, and 20 percent were White). The researchers noted that previous research had demonstrated stereotype threat effects for Black, Hispanic, and low-income students and argued that, for this reason, “all of the participants in the sample were potentially susceptible to stereotype threat” (p. 652). In the Cohen, Garcia, Apfel, and Master (2006) study, participants were from a suburban northeastern middle school with a student population equally split between Black and White students. Whereas the other two studies were conducted in socioeconomically disadvantaged settings, this study was conducted in a suburban area. However, race (Black or White) was used as a factor in the analyses (119 Black and 124 White students participating). Interestingly, all three included studies focused on grade 7 students.

Five studies were excluded because they did not include K–12 students as their focus. Though the studies examined the impact of an intervention on improving Black student performance, the sample was college students in laboratory settings, not K–12 students. Thus, these studies lacked external validity. Although a common practice in certain disciplines, it is difficult to generalize results from studies conducted with college students to other populations, especially to populations that are significantly younger.

---

### Verification search

Because of the small number of studies identified for inclusion, a second broader, verification search was conducted to catch any relevant studies that might have been missed in the focused search of the databases. This verification search used the broadest search term of “stereotype threat” without the word “intervention,” searching the literature using the terms “stereotype threat,” “stereotype,” and “threat.” The EBSCO host search engine was used to search the ERIC, PsycINFO, Academic Search Premier, and Soc Index with Full Text databases. Also, the Education Index database was searched using Wilson Web, and the Dissertations Abstracts database was searched using First Search. The entire text of identified documents was searched, not just keywords or title. The only limit placed on the search was the publication year, which was set at between 1990 and 2007 (as the concept of stereotype threat emerged in the 1990s).

This search identified 741 references. Reviews of the titles and abstracts turned up no additional studies appropriate for inclusion. The reasons for exclusion were as follows: 74 percent were off-topic, 14 percent were not empirical, and 12 percent were on-topic but did not test an intervention, occur in K–12 classrooms, or include Black students.

---

### Review process: identifying methodological limitations of included studies

The three studies identified as meeting the six inclusion criteria in the article screening protocol

(in appendix C) were reviewed first by a Regional Educational Laboratory (REL) Southeast researcher using a study quality review protocol (see appendix C). The researcher adapted the items on the protocol from one used by REL Central, which provided the researcher with background knowledge about the meaning of each item. The completed protocols for each study and the study articles were then examined by an external reviewer trained in What Works Clearinghouse (WWC) criteria.

### *Development of study quality review protocol.*

Researchers for this study obtained a copy of a coding protocol that REL Central had developed using the WWC evidence standards (U.S. Department of Education 2008) to code studies included in the report *Using strategy instruction to help struggling high schoolers understand what they read* (Apthorp and Clark 2007). This coding protocol included criteria that WWC indicates are important, such as adequacy of outcome measure, equivalence of groups at baseline, extent of overall and differential group attrition, intervention contamination, and confounding of teacher and intervention. Also included were descriptive items to summarize each study, such as independent and dependent variable description, summary of analysis and results, and an overall narrative summary of the study.

The REL Central coding protocol was simplified for this study, as the intention was to describe any limitations in the methodology of the three studies based on an interpretation of WWC standards and the researchers’ understanding of good science, rather than to conduct a WWC-level review. The REL Southeast staff member who developed the protocol and who has experience in research design used the study quality review protocol to gather information from each study on items in the protocol: adequacy of outcome measure, random assignment process, overall attrition, differential group attrition, intervention contamination, and confounding factors. A section was not included on items related to assessing the quality of quasi-experimental designs in the protocol since all three identified studies used an experimental design. The completed coding protocol on each study was

reviewed by the external reviewer, who raised questions for clarification with the third researcher from REL Southeast and the initial coder.

*Assessing the quality of identified intervention studies.* The three studies were subject to a final quality review to describe any methodological limitations, using a study coding protocol (see appendix C) based on the five criteria below from the *What Works Clearinghouse Procedures and Standards Handbook* (U.S. Department of Education 2008) for assessing the internal validity of studies examining the effects of interventions:

- *Outcome measures.* The measures used to assess impact must be shown to actually measure what they are intended to measure. For studies in school settings, common academic achievement measures include state- or locally mandated tests and course performance (term grades). The three studies reported on here used such school measures of student achievement.
- *Random assignment process.* In experimental studies researchers use random assignment to assign participants to experimental conditions (intervention or control) to ensure that the groups are as similar as possible on all characteristics so that the outcomes measured reflect the influence of the intervention only. All three of the studies reported on their random assignment process, so any threats to random assignment could be identified. Only one study had a limitation in this area (Good, Aronson, and Inzlicht 2003).
- *Attrition of participants.* Loss of participants can create differences in measured outcomes by changing the composition of the intervention or control groups. Both overall attrition and differential attrition (differences between intervention and control groups) are of concern. All three studies were acceptable in this area.
- *Intervention contamination.* Intervention contamination can happen when unintended events occur after intervention begins.

Because these new factors could affect group outcomes, they also could affect the conclusions of the experiment. An example is a teacher in an intervention group sharing the intervention materials with a teacher in a control group. One study was noted as having a possible limitation in this area (Good, Aronson, and Inzlicht 2003).

- *Confounding factor.* It is important to examine factors beyond the intervention that might affect differences between groups, such as the effects of teachers or of the intervention provider more generally. For example, if each condition of the study involves only one teacher's classroom, then the effects of the teacher cannot be separated from the effects of the intervention. No studies were noted as having problems in this area.

*Methodological review.* The methodological limitations reported for each study were identified through this process. The results of the study quality review process are shown in the individual descriptions of each study below and summarized in table B3. Table B4 summarizes the methodology of the three studies.

*Blackwell, Trzesniewski, and Dweck (2007).* No limitations were noted in applying the quality review criteria to this study:

- *Random assignment process.* Students were randomly assigned by the school to regularly scheduled advisory classes (groups of 12–14). Each pre-existing advisory group was assigned by the research team to an intervention or control condition. The researchers reported baseline equivalence data: fall term math grades for the students were not significantly different for the two groups (2.38 for the intervention group and 2.41 for the control group).
- *Attrition.* The attrition rate (students who did not complete the eight-week sessions) was 5 percent and roughly equivalent for both groups (three from the intervention group and two from the control group).

TABLE B3  
Quality of final studies included in report

Study	Adequacy of outcome measure			Random assignment process			Overall attrition			Differential attrition			Intervention contamination		Confounding factor	
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	1	2
Blackwell, Trzesniewski, and Dweck (2007)	✓			✓			✓			✓			✓		✓	
Good, Aronson, and Inzlicht (2003)	✓				✓		✓			✓				✓	✓	
Cohen, Garcia, Apfel, and Master (2006)	✓			✓			✓			✓			✓		✓	

Note: 1 = acceptable; 2 = acceptable with reservations; 3 = not acceptable.

Source: Authors' compilation.

- *Intervention contamination.* There was no reporting of any events during the eight weekly 25-minute periods that might differentially affect the two groups. Each advisory group was assigned to a condition, making it less likely students would share information across conditions.
- *Confounding factors.* The study used undergraduate assistants to deliver the eight sessions, assigning two undergraduates as workshop leaders for each advisory class. Different workshop leaders were assigned to each advisory class. Student participants all had the same math teacher during the study period, so differences in math teachers could not have influenced differences in math grades between the intervention and control students.

*Good, Aronson, and Inzlicht (2003).* Two limitations were noted in applying the study quality review criteria that might limit the confidence in the results of this study.

- *Random assignment process.* Six of the 138 students' scores were removed from the analysis. In addition, evidence was not presented on the equivalence of the four groups on baseline

achievement. Although the authors reported that the six excluded students did not come from any particular condition, it is difficult to know how well the random assignment process worked in creating equivalent groups at baseline. Therefore, results showing differences between experimental conditions after the intervention should be interpreted with caution.

- *Intervention contamination.* The same mentors provided the intervention to students in three of the four experimental conditions, so the intervention conditions could have been somewhat blurred if the mentors brought knowledge from one condition to their delivery of another. In addition, students were all in the same class so they could have discussed or shared their experiences across the experimental conditions. Such a problem would work against finding a significant difference between the control group and the other experimental conditions, thus, perhaps strengthening confidence in the intervention condition effects where found. (Under WWC review standards, contamination such as occurred in this study is not considered grounds for downgrading a study.)



No limitations were found relative to the attrition or confounding factors.

- *Attrition.* Roughly 4 percent of students were excluded from the reading test analysis based on an outlier analysis intended to identify students whose test score results represented very limited English speaking skills. This attrition rate is less than the 20 percent level determined as significant attrition. The attrition was reported as occurring equivalently across groups.
- *Confounding factors.* The participating students were all part of one class, but the teacher did not provide the intervention. Students were randomly assigned to one of four conditions and also randomly assigned to a mentor who provided their condition.

*Cohen et al. (2006).* No limitations were noted in applying the quality review criteria to this study, as summarized below.

- *Random assignment process.* The article reported on two randomized, double-blind experiments of an affirmation intervention. Students in three teachers' classrooms were involved. Random assignment to either the affirmation intervention or control condition was at the level of the individual student. For each teacher/classroom period, there were about equal numbers of students in the two conditions. Baseline measures for each student (standardized measure of pre-intervention in-class performance, prior year grade point average in core courses, and pre-intervention test score) were collected and used in the analysis as potential covariates.
- *Attrition.* Individual student attrition (absences, missing data, experimenter error) was four students for study 1 (roughly 3 percent attrition), leaving 111 students in the final sample, and seven students from study 2 (roughly 5 percent), leaving 132 students in the final sample. There was no differential attrition as a function of condition, as indicated by the authors in a subsequent correspondence; baseline covariates were used in the analysis.
- *Intervention contamination.* There was no reporting of events or circumstances that might have contributed to contamination. The experiment was double-blind, so the teachers did not know what condition the students were assigned to, nor did the students. Additionally, neither group was aware of the experimental hypothesis, and students were unaware of the intervention.
- *Confounding factors.* Students were the unit of analysis for the study and were randomly assigned to the two conditions in approximately equal numbers for each of the three teachers. Because fall grades in the targeted course were the outcome measure and teachers may grade differently, the regression analysis included a teacher variable (dummy codes for the three teachers), a main effect of baseline in-class performance measures, and two terms representing the interaction of baseline in-class performance with each of the two teacher dummy variables to control for teacher differences in the predictiveness of early in-class performance. Thus, teacher effects were addressed and did not threaten internal validity.

TABLE B4

**Methodological summary of the three experimental studies reviewed**

Study	Intervention description	Sample	Design	Measures and outcomes	Analysis methods	Findings
Blackwell, Trzesniewski, and Dweck (2007) (Study 2)	Students participated in eight weekly 25-minute sessions. During the sessions students engaged in reading, activities, and discussions related to the malleability of intelligence. Students in the control condition also received eight sessions, but they received materials on the brain and memory for four of the sessions rather than the intervention materials on how to grow intelligence.	91 students in grade 7 in a large urban school district (48 in the experimental condition and 43 in the control condition): <ul style="list-style-type: none"> <li>• 52 percent were Black.</li> <li>• 45 percent were Hispanic.</li> </ul>	Random assignment by advisory class periods to two conditions: intervention and control. Both groups received eight weeks of sessions, but the control group sessions did not include the information on the malleability of intelligence.	1. Theory of intelligence questionnaire 2. Changes in classroom motivation 3. Grade 7, spring term scores in math (time 3)	1. Analysis of variance 2. Chi-square test 3. Hierarchical linear modeling	<p>Students in the intervention group showed a significantly greater change from pretest to posttest on the theory of intelligence questionnaire than did the control group and scored significantly higher on the measure after the intervention than did the control group (<math>d = 0.47</math>), indicating that the theory of intelligence message was successfully communicated.</p> <p>2. Teachers identified 27 percent of the intervention group as showing positive motivational change, compared with 9 percent in the control group, a significant difference.</p> <p>3. "There was a significant effect of experimental condition on change in grades across the intervention (Time 2 to Time 3; <math>b = 53, t = 2.93, p &lt; .05</math>). Thus, the sample as a whole was decreasing in grades, but this decline was eliminated for those in the experimental condition. The decline in grades suffered by the control-group students mirrors that commonly observed over the junior high school transition" (p. 257).</p>

(CONTINUED)

TABLE B4 (CONTINUED)

**Methodological summary of the three experimental studies reviewed**

Study	Intervention description	Sample	Design	Measures and outcomes	Analysis methods	Findings
Good, Aronson, and Inzlicht (2003)	<p>Students were randomly assigned mentors (students from a local college trained to deliver the intervention) who taught them messages throughout the school year through personal and email communication. There were three interventions that were hypothesized to help students deal with stereotype threat: students were taught that intelligence is malleable, school difficulty is normal, or both. Mentors for the control group students taught them about the dangers of drugs (a topic unrelated to the interventions).</p>	<p>138 grade 7 students who were enrolled in a computer skills class in a rural district in Texas, 70 percent of them eligible for free or reduced-price lunch:</p> <ul style="list-style-type: none"> <li>• 67 percent of sample were Hispanic.</li> <li>• 13 percent were Black.</li> <li>• 20 percent were White.</li> </ul> <p>The researchers note that previous research demonstrated negative stereotype threat effects on performance for Black students, Hispanic students, girls in math, and students from low-income households. Thus, “all of the participants in the sample were potentially susceptible to stereotype threat” (p. 652).</p>	<p>Random assignment of students in the computer skills course to one of four conditions: incremental intelligence, attribution of difficulty, combination, and antidrug (control). The mentors were blind to the specific hypotheses of the study.</p>	<p>Achievement scores on statewide standardized tests of reading and mathematics (Texas Assessment of Academic Skills)</p>	<p>1. For math, a 2 (gender) by 4 (experimental condition) analysis of variance (ANOVA) was conducted followed by planned comparisons. 2. For reading, test scores were submitted to a one-way ANOVA comparing performance across the four conditions followed by planned comparisons. 3. In both sets of analyses, the sample size was too small to examine differences between ethnic groups.</p>	<p>1. Math results indicated a significant main effect of gender, qualified by a significant gender-condition interaction; girls in the incremental (<math>d = 1.13</math>), attributional (<math>d = 1.50</math>), or combined (<math>d = 1.30</math>) condition achieved significantly higher math scores than girls in the control condition. “All these are large effect sizes indicating that the intervention procedures meaningfully increased females’ math scores compared to the control condition” (p. 656). Marginally significant improvement was found for boys in the incremental condition compared with boys in the control condition (<math>d = 0.64</math>). 2. Reading. The ANOVA revealed a significant effect of condition. Planned comparisons indicated that students in the incremental (<math>d = 0.52</math>) and attributional (<math>d = 0.71</math>) conditions achieved significantly higher reading scores than students in the control condition.</p>

(CONTINUED)

TABLE B4 (CONTINUED)

**Methodological summary of the three experimental studies reviewed**

Study	Intervention description	Sample	Design	Measures and outcomes	Analysis methods	Findings
Cohen, Garcia, Apfel, and Master (2006)	The intervention was a self-affirmation exercise in which students completed an in-class writing assignment early in the school year, writing about an important value they held. The writing exercise (provided once in study 1 and twice in the replication study 2) was provided without teacher instruction. Students read instructions provided in an envelope, completed the instructions, placed their work in the envelope and returned it to the teacher. In the affirmation treatment, students were asked to select the most important personal value (or values) from a list and asked to write a paragraph about why the value was important to them. Control group students were asked to select the least important value and to write about why the value might be important to someone else.	Participants were seventh graders in a suburban northeastern middle school with a student body split evenly between Black and White students. 243 students participated in the two studies reported on (study 2 was a replication of study 1 with different students): <ul style="list-style-type: none"> <li>• 119 were Black.</li> <li>• 124 were White.</li> </ul> The treatment group had 119 students, and the control group had 124 students.	Students were randomly assigned to one of two conditions: self-affirmation treatment or control condition, resulting in four cells: race (Black or White) by 2 conditions (affirmation or control writing assignment)	Fall semester grade point average in the targeted academic course. The intervention was delivered early in the fall semester in the targeted course. The targeted course was the same subject area for all experimental and control students.	1. Multiple regression	1. Black students participating in the self-affirmation condition earned a higher grade in the targeted course than students in the control group (average difference between intervention and control group grades was 0.30 grade point). No difference was found between the conditions for White students. The race-condition interaction was significant for both studies.  2. Combining data from both studies showed that Black students participating in the intervention were less likely to receive a D or below in the course (9 percent) than were students in the control condition (20 percent).

Source: Authors' compilation.

## APPENDIX C ARTICLE SCREENING AND STUDY QUALITY REVIEW PROTOCOLS

### Article screening protocol

Coder (name and date): \_\_\_\_\_

APA style citation: \_\_\_\_\_

### Initial-level screening for relevance

#### Content relevance

1. Is the article on topic?
  - Yes
  - No (exclude)
2. Is the citation an empirical study (not a literature review, book chapter, conceptual paper, etc.)?
  - Yes
  - No (exclude)
3. Is the study focused on race-based stereotype threat (as opposed to gender or other type of stereotype threat)?
  - Yes
  - No (exclude)

### Second-level screening for relevance

#### Intervention-type relevance

4. Does the study investigate interventions aimed at reducing stereotype threat?
  - Yes
  - No (exclude)

#### Sample relevance

5. Does the study include African American students?
  - Yes
  - No (exclude)
6. Does the study focus on K–12 students?
  - Yes
  - No (exclude)

*Relevance screen summary.* In order for the study to be included, it must have passed all relevance screens (content relevance, intervention-type relevance, and sample relevance). If “yes” for all six items above, the study is eligible for inclusion in the report if it is judged of sufficient methodological quality applying What Works Clearinghouse–based criteria.

### Study quality review protocol

### Study information and outcome measure

#### 1. Study information

- a. Reference citation (author and publication year):  
\_\_\_\_\_
- b. Title: \_\_\_\_\_
- c. Source:
  - Dissertation
  - Conference presentation
  - Technical report
  - Book or book chapter
  - Journal (specify name): \_\_\_\_\_

2. *Adequacy of outcome measure.* List the outcome measures and the validity and reliability evidence as outlined below.

Examples of validity evidence includes test or measure has correlations from studies of concurrent validity, predictive validity, factor analysis; measure is in established use as an academic achievement indicator (for example, a state-developed standardized test administered as part of an annual student testing program or course grades on official transcripts) and thus has face validity as a reliable measure of student achievement. Examples of reliability evidence include internal consistency, test-retest, or, if measure requires judgment, interrater reliability.

### Quality review criteria

3. *Random assignment process.* In looking at information included in the study on the random assignment process . . . (check one)

- There were no reported disruptions of or contaminations in random assignment process, and/or baseline equivalence was checked. → (Acceptable)
  - There was evidence of disruptions or contaminations in random assignment process, but they were minor and/or pre-test differences were checked and were nonsignificant → (Acceptable with reservations)
  - There was evidence of disruptions or contaminations in random assignment process, and pretest differences were not checked or were checked, were significant, and were not corrected statistically. → (Not acceptable)
4. *Attrition*. In looking at information included in the study on attrition . . . (check one)
- There was no significant attrition (<20 percent overall). → (Acceptable)
  - There was significant attrition (>20 percent overall), but postattrition equivalence was demonstrated. → (Acceptable)
  - There was significant attrition (>20 percent overall), and postattrition equivalence was not demonstrated. → (Acceptable with reservations)
  - There was no information on attrition provided, but degrees of freedom provide adequate information and indicate no significant attrition (<20 percent overall). → (Acceptable)
  - There was no information on attrition provided, but degrees of freedom provide information that indicates significant attrition (>20 percent overall), and postattrition equivalence was demonstrated. → (Acceptable)
  - There was no information on attrition provided, but degrees of freedom provide information that indicate significant attrition (>20 percent overall), and postattrition equivalence was not demonstrated. → (Acceptable with reservations)
- There was no information on attrition provided, and degrees of freedom do not provide adequate information. → (Not acceptable)
5. *Differential sample attrition*. In looking at information included in the study on sample attrition . . . (check one)
- There was no significant attrition differential between intervention and comparison groups (<7 percent). → (Acceptable)
  - There was significant attrition differential (>7 percent), but group comparability was demonstrated. → (Acceptable)
  - There was significant attrition differential (>7 percent), and group comparability was not demonstrated. → (Acceptable with reservations)
  - There was no information on attrition differential provided, but degrees of freedom provide adequate information and indicate no significant attrition (<7 percent). → (Acceptable)
  - There was no information on attrition differential provided, but degrees of freedom provide information that indicate significant attrition (>7 percent); however, group comparability was demonstrated. → (Acceptable)
  - There was no information on attrition differential provided, but degrees of freedom provide information that indicate significant attrition (>7 percent); group comparability was not demonstrated. → (Acceptable with reservations)
  - There was no information on attrition differential provided, and degrees of freedom do not provide adequate information. → (Not acceptable)
6. *Intervention contamination*. Was there evidence of something happening after the beginning of the intervention that affects the outcomes for the intervention or control group (affects the outcome of one of the groups in an unexpected way)? (check one)
- No → (Acceptable)
  - Yes → (Acceptable with reservations)

7. *Confounding factor* (teacher or other intervention delivery agent confounds). In looking at information included in the study on the assignment or role of teachers or other intervention delivery agents (e.g., mentors), was there any situation in which one teacher or other intervention delivery agent was assigned to just one experimental condition? (check one)
- No → (Acceptable)
  - Yes → (Not acceptable)

8. *Summary of randomized controlled trial study quality review criteria.* Look back at your answers for questions 2–7 and enter the results in the following box:

Criteria	Acceptable with reservations		
	Acceptable	Acceptable with reservations	Not acceptable
Q2. Adequacy of outcome measure			
Q3. Random assignment process			
Q4. Overall attrition			
Q5. Differential sample attrition			
Q6. Intervention contamination			
Q7. Confounding factor			

*Overall study description.* The following items are used to ensure that consistent information was gathered about each study.

9. *Study population sample.*

- a. School district/local:
- Urban
  - Suburban
  - Rural
  - Missing

- b. Race/ethnicity of students included (check all that apply):
- African American/Black
  - Asian/Pacific Islander
  - Hispanic/Latino
  - White
  - Other
  - Multiracial
- c. Percentage of students receiving free or reduced-price lunch: \_\_\_\_\_
- d. Grade level (check all that apply):
- K–3
  - 3–5
  - 6–8
  - 9–12
- e. Age (mean and/or minimum–maximum): \_\_\_\_\_
- f. Total sample size: \_\_\_\_\_
- g. Achievement outcome variable (if more than one treatment group, specify in parentheses which treatment group you are placing in each column)

Outcome	Achievement outcome variable	Standard score mean
Outcome 1 (specify)	Treatment 1 (specify)	
	Treatment 2 (specify)	
	Treatment 3 (specify)	
	Control	
	Overall	
Outcome 2 (specify)	Treatment 1 (specify)	
	Treatment 2 (specify)	
	Treatment 3 (specify)	
	Control	
	Overall	

*Independent variable/intervention*

10. *Intervention description*

- a. Briefly describe the intervention, including the stated purpose and any required special conditions or resources: \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_
- b. Subject area (check all that apply):  
 English/language arts  
 Math  
 Social studies  
 Science  
 Other (specify course title)  
 \_\_\_\_\_
- c. Duration of intervention: \_\_\_\_\_

*Analysis and results.*

11. Unit of assignment and analysis match.

- a. Was there a match between unit of assignment and analysis? (Check one)  
 Matched, both were students  
 Matched, both were teachers  
 Matched, both were schools  
 Not matched, not addressed in analyses, group differences not statistically significant  
 Not matched, not addressed in analyses, group differences statistically significant  
 Not matched, but addressed in analyses  
 Explain: \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_
- b. Was an effect size reported?  
 No  
 Yes (specify pages) \_\_\_\_\_

12. *Results.* Please fill in the following table for each outcome included in the study

Outcome	Statistic	Notes	Page numbers
Outcome 1 (specify)	Mean/count/proportions (include both treatment and control statistics)		
	Sample size (include both treatment and control statistics)		
	Standard deviation (include both treatment and control statistics)		
	Test statistic		
	Were group differences statistically significant? (provide p-value)		
	Researcher reported effect size (including type, if available)		
	Outcome 2 (specify)	Mean/count/proportions (include both treatment and control statistics)	
Outcome 2 (specify)	Sample size (include both treatment and control statistics)		
	Standard deviation (include both treatment and control statistics)		
	Test statistic		
	Were groups differences statistically significant (provide p-value)?		
	Researcher reported effect size (including type, if available)		



## REFERENCES

- Apthorp, H., and Clark, T. (2007). *Using strategy instruction to help struggling high schoolers understand what they read* (Issues & Answers Report, REL 2007–No. 038). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central.
- Aronson, J. (2002). Stereotype threat: contending and coping with unnerving expectations. In J. Aronson (Ed.), *Improving academic achievement: impact of psychological factors on education*. San Diego, CA: Academic Press.
- Aronson, J., Fried, C., and 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, 113–25.
- Aronson, J., and Good, C. (2002). The development and consequences of stereotype vulnerability in adolescents. In F. Pajares and T. Urdan (Eds.), *Adolescence and education*. New York: Information Age Publishing.
- Aronson, J., and Inzlicht, M. (2004). The ups and downs of attributional ambiguity: stereotype vulnerability and the academic self-knowledge of African American students. *Psychological Science*, 15(12), 829–36.
- Aronson, J., Lustina, M.J., Good, C., Keough, K., Steele, C.M., and Brown, J. (1999). When white men can't do math: necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, 35(1), 29–46.
- Aronson, J., and Steele, C.M. (2005). Stereotypes and the fragility of human competence, motivation, and self-concept. In C. Dweck and E. Elliot (Eds.), *Handbook of competence and motivation*. New York: Guilford.
- Beilock, S.L., Jellison, W.A., McConnell, A.R., and Carr, T.H. (2006). On the causal mechanisms of stereotype threat: can skills that don't rely heavily on working memory still be threatened? *Personality and Social Psychology Bulletin*, 32(8), 1059–71.
- Blackwell, L., Trzesniewski, K., and Dweck, C.S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: a longitudinal study and an intervention. *Child Development*, 78(1), 246–63.
- Bowen, W.G., and Bok, D. (1998). *The shape of the river: Long-term consequences of considering race in college and university admissions*. Princeton, NJ: Princeton University Press.
- Brown, R.P., and Lee, M.N. (2005). Stigma consciousness and the race gap in college academic achievement. *Self & Identity*, 4, 149–57.
- Chiu, C.Y., Hong, Y.Y., and Dweck, C.S. (1997). Lay dispositionism and implicit theories of personality. *Journal of Personality and Social Psychology*, 73(1), 19–30.
- Cohen, G.L., Garcia, J., Apfel, N., and Master, A. (2006). Reducing the racial achievement gap: a social-psychological intervention. *Science Magazine*, 313(5791), 1307–10.
- Cohen, G.L., Steele, C.M., and Ross, L.D. (1999). The mentor's dilemma: providing critical feedback across the racial divide. *Personality and Social Psychology Bulletin*, 25(10), 1302–18.
- Creswell, J.D., Welch, W.T, Taylor, S.E., Sherman, D.K, Gruenewald, T.L., and Mann T. (2005). Affirmation of personal values buffers neuroendocrine and psychological stress responses. *Psychological Science*, 16(11), 846–51.
- Croizet, J., and Claire, T. (1998). Extending the concept of stereotype threat to social class: the intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, 24(6), 588–94.
- Dweck, C.S. (1999). Caution—praise can be dangerous. *American Educator*, 23(1), 4–9.
- Good, C., Aronson, J., and Inzlicht, M. (2003). Improving adolescents' standardized test performance: an intervention to reduce the effects of stereotype threat.

- Journal of Applied Developmental Psychology*, 24(6), 645–62.
- Hacker, A. (1995). *Two nations: Black and White, separate, hostile, unequal*. New York: Ballantine.
- Inzlicht, M., and 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, 365–71.
- Jencks, C., and Phillips, M. (1998). *The Black-White test score gap*. Washington, DC: The Brookings Institution.
- Johns, M., Schmader, T., and Martens, A. (2005). Knowing is half the battle: teaching stereotype threat as a means of improving women's math performance. *Psychological Science*, 16(3), 175–79.
- KewalRamani, A., Gilbertson, L., Fox, M., and Provasnik, S. (2007). *Status and trends in the education of racial and ethnic minorities* (NCES 2007-039). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center of Education Statistics.
- Marx, D.M., and Roman, J.S. (2002). Female role models: protecting women's math test performance. *Personality and Social Psychology Bulletin*, 28(9), 1183–93.
- McGlone, M., and Aronson, J. (2006). Social identity salience and stereotype threat. *Journal of Applied Developmental Psychology*, 27(5), 486–93.
- McGlone, M.S., and Aronson, J. (2007). Forewarning and forearmng stereotype-threatened students. *Communication Education*, 56(2), 119–33.
- Mendoza-Denton, R., Downey, G., Purdie, V.J., Davis, A., and Pietrzak, J. (2002). Sensitivity to status-based rejection: implications for African American students' college experience. *Journal of Personality and Social Psychology*, 83(44), 896–918.
- Quinn, D.M., and Spencer, S.J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues*, 57(1), 55–71.
- Rosenthal, H., and Crisp, R.J. (2006). Reducing stereotype threat by blurring intergroup boundaries. *Personality and Social Psychology Bulletin*, 32(4), 501–11.
- Rothstein, R. (2002). *Classes and schools: using social, economic, and educational reform to close the Black-White achievement gap*. Washington, DC: Economic Policy Institute.
- Ryan, K.E., and Ryan, A.M. (2005). Psychological processes underlying stereotype threat and standardized math test performance. *Educational Psychologist*, 40(1), 53–63.
- Schmader, T., and Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85(3), 440–52.
- Shapiro, J.R., and Neuberg, S.L. (2007). From stereotype threat to stereotype threats: implications of a multi-threat framework for causes, moderators, mediators, consequences, and interventions. *Personality and Social Psychology Review*, 11(2), 107–30.
- Southern Regional Education Board. (2003). *ACT and SAT scores in the south: the challenge to lead*. Retrieved December 10, 2007, from [www.sreb.org/main/highschools/college/ACT-SAT\\_2003\\_.pdf](http://www.sreb.org/main/highschools/college/ACT-SAT_2003_.pdf).
- Spencer, S.J., Steele, C.M., and Quinn, D.M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4–28.
- Steele, C.M. (1988). The psychology of self-affirmation: sustaining the integrity of the self. *Advances in Experimental Social Psychology*, 21, 261–302.
- Steele, C.M. (1997). A threat in the air: how stereotypes shape the intellectual identities and performance of women and African Americans. *American Psychologist*, 52(6), 613–29.
- Steele, C.M., and 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., and 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. 37. San Diego, CA: Academic Press.
- Stone, J., Lynch, C.I., Sjomeling, M., and Darley, J.M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology*, 77(6), 1213–27.
- U.S. Department of Education. 2008. *WWC Procedures and Standards Handbook. Version 2.0*—December. Washington, DC: U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. Retrieved December 10, 2007, from <http://ies.ed.gov/ncee/wwc/references/idocviewer/doc.aspx?docid=19&tocid=1>.
- Wald, J., and Losen, D. (2005, May 17). *Confronting the graduation rate crisis in the south*. The Civil Rights Project, Harvard University, Cambridge, MA.
- Walton, G.M., and Cohen, G.L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, 39(5), 456–67.
- Walton, G.M., and Cohen, G.L. (2007). A question of belonging: race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92(1), 82–96.
- Wheeler, S.C., and Petty, R.E. (2001). The effects of stereotype activation on behavior: a review of possible mechanism. *Psychological Bulletin*, 127(6), 797–826.
- Whimbey, A. (1975). *Intelligence can be taught*. New York: Dutton.
- Wilson, T.D., and Linville, P.W. (1985). Improving the performance of college freshmen with attributional techniques. *Journal of Personality and Social Psychology*, 49, 287–93.