

Stereotype threat and test performance: A primer for school psychologists

Alexander H. Jordan ^{a,*}, Benjamin J. Lovett ^b

^a *Department of Psychology, Stanford University, Jordan Hall, Bldg. 420, Stanford, CA 94305, United States*

^b *Department of Psychology, Syracuse University, 430 Huntington Hall, Syracuse, NY 13244, United States*

Received 9 March 2006; received in revised form 5 August 2006; accepted 5 September 2006

Abstract

Ethical guidelines require school psychologists to ensure that their assessment practices are nondiscriminatory, but typical discussions on this topic neglect the possible discriminatory effects of cultural stereotypes on assessment results. Recent research on *stereotype threat* shows that students' knowledge of stereotype-based negative expectations about their test performance can depress their actual test performance. This paper discusses the range of conditions that promote stereotype threat and identifies important moderators and mediators of the phenomenon. Several practical suggestions are offered for school psychologists to consider when interviewing students, interpreting assessment results, and developing programs to increase schoolwide achievement.

© 2006 Society for the Study of School Psychology. Published by Elsevier Ltd. All rights reserved.

Keywords: Stereotype threat; Stereotypes; Academic performance; Psychological assessment; School psychology

For almost a century, critics have voiced concern about bias and discrimination in psychological assessment, beginning with the administration of standardized cognitive ability tests to immigrants who did not speak English and soldiers who could not read (Zenderland, 1998). Persistent claims of bias in test questions and scoring procedures have even led to occasional calls for the abandonment of all standardized testing (e.g., Williams, 1971). In response to such criticism, professional organizations have issued ethical guidelines mandating that psychologists conduct assessments and interpret results in a fair

* Corresponding author. Tel.: +1 650 497 1337.

E-mail addresses: alex.jordan@stanford.edu (A.H. Jordan), bjlovett@syr.edu (B.J. Lovett).

and nondiscriminatory manner (e.g., American Psychological Association, 2002; National Association of School Psychologists, 2000). These general prescriptions have led, in turn, to articles, chapters, and entire books devoted to training psychologists in a variety of specific nondiscriminatory assessment practices (e.g., Ortiz, 2002; Padilla, 2001; Valencia & Suzuki, 2000).

Nearly absent from the growing literature on culturally sensitive assessment (e.g., Lopez & Rogers, 2001) has been discussion of recent social psychological research on *stereotype threat*. According to this research, largely conducted using college student samples, negative cultural stereotypes about the performance of certain demographic groups can actually depress the performance of examinees belonging to those groups. Despite the obvious relevance of this phenomenon to nondiscriminatory assessment practices, only Frisby (1999) has mentioned stereotype threat in a school psychology context. After providing a brief critical discussion of a single empirical study (Steele & Aronson, 1995), Frisby (1999) concluded that although the concept of stereotype threat was provocative, far more research needed to be done before any broad implications for school psychologists could be deduced.

In the years since Frisby's article, a large literature on stereotype threat has accumulated. The goal of the present paper is to provide an up-to-date review of this literature and to enumerate several implications for the practicing school psychologist. A working knowledge of stereotype threat is especially important for school psychologists in today's assessment milieu, where standardized tests that were once viewed as context-free measurements of students' abilities, skills, and personalities are now recognized as limited snapshots of individuals' behaviors at a single point in time, susceptible to influence by immediate and long-term environmental factors (see, e.g., the "intelligent testing" philosophy, Kaufman, 1994, 2000; and the behavioral-systems approach, Mash & Terdal, 1997).

What is stereotype threat?

All people sometimes experience situations in which, according to cultural stereotypes, their social identities augur poorly for their performance. Awareness of these negative stereotypes may produce distracting thoughts about confirming group stereotypes, and these anxieties, in turn, may lead to the very failure that is feared. For example, a girl who tries out for an improvisational comedy troupe in high school, aware of a stereotype that women do not make good comedians, may find herself flustered and unable to perform to her potential. She may, furthermore, attribute her comedic stumbles to her *femaleness* and become dejected in a way that male members of her troupe do not when they experience frustration. According to the theory and terminology developed by social psychologist Claude Steele (1997; Steele, Spencer, & Aronson, 2002), this girl's experience exemplifies stereotype threat.

Stereotype threat's performance-depressing effects have been demonstrated in a variety of domains. In one study, for example, White men performed poorly compared to Black men when given a miniature golf task described as a test of natural athletic ability, consistent with the stereotype that White individuals are poor natural athletes (Stone, Lynch, Sjomeling, & Darley, 1999). When the same task was described as a test of "sporting intelligence," on the other hand, where the cultural stereotype is reversed, Black

men's performance suffered. More relevant to school psychologists, many well-designed laboratory studies have shown stereotype threat effects on the academic test performance of vulnerable demographic group members.

Stereotype threat and academic testing

Ethnic minority students and test performance

In the first empirical study examining stereotype threat, [Steele and Aronson \(1995\)](#) administered difficult verbal GRE items to Black and White college students. Half of the students were told that the test measured verbal ability, whereas the other half were told that the test had nothing to do with ability. The experimenters hypothesized that the former condition would induce stereotype threat in the Black students and hurt their performance, given the cultural stereotype that Black individuals, relative to White individuals, lack intellectual ability (e.g., [Ashton & Esses, 1999](#)). In the latter condition, on the other hand, it was expected that this stereotype would not be salient to those taking the test, and consequently Black and White students would do equally well.

The results of the study were consistent with the experimenters' hypotheses. When told that the test was diagnostic of verbal ability, Black students scored a full standard deviation lower than the White students. But when the test was said to be nondiagnostic of ability, the Black and White students performed equally well. It seemed, therefore, that the situational relevance of a racial stereotype had a major impact on test performance. Indeed, in a second study, a performance difference between Black and White students was produced simply by having students specify their race on a demographics form before testing, even when the test was represented as nondiagnostic of ability ([Steele & Aronson, 1995](#)). Similar effects have been obtained using different university populations (e.g., [Blascovich, Spencer, Quinn, & Steele, 2001](#)) and even when using the Raven Progressive Matrices, generally considered to be a "culture free" measure of intelligence ([McKay, Doverspike, Bowen-Hilton, & Martin, 2002](#)).

Two means of evoking cultural stereotypes, then, have been shown in laboratory studies to produce stereotype-consistent subsequent test performance in Black college students: (1) designating a test as diagnostic of intellectual ability, and (2) bringing attention to students' ethnicities prior to testing. Similar patterns of results have been obtained for Latino college students (e.g., [Gonzales, Blanton, & Williams, 2002](#)), who also suffer negative stereotypes in the academic domain, as well as low-SES students, who are negatively stereotyped in verbal fluency ([Croizet & Claire, 1998](#)).

Female students and test performance

Female students from elementary through graduate school are stereotyped as deficient in mathematics ([Swim, 1994](#)). We might expect, therefore, that they would be vulnerable to stereotype threat when completing mathematically intensive tasks. Indeed, in one study, female college students performed worse than their male counterparts on a very difficult math test, but this performance difference was eliminated when the test was represented to the participants as not showing gender differences ([Spencer, Steele, & Quinn, 1999](#)).

Presumably, this manipulation reduced stereotype threat by making the women's gender irrelevant to the task at hand—just as describing a test as nondiagnostic of ability can render Blacks' ethnicity irrelevant to the task at hand. Stereotype threat's detrimental effects on female students' math performance have been demonstrated using several other distinct methods to activate the relevant negative stereotypes (e.g., Brown & Josephs, 1999; Cadinu, Maass, Rosabianca, & Kiesner, 2005; Davies, Spencer, Quinn, & Gerhardstein, 2002; Inzlicht & Ben-Zeev, 2000).

The mechanics of stereotype threat

Physiological arousal

Stereotype activation in the mind of the test-taker seems crucial to stereotype threat (e.g., Davies, Spencer, & Steele, 2005; Steele & Aronson, 1995), but it is unclear exactly how the activation of self-relevant negative stereotypes affects test performance (see Major & O'Brien, 2005, for a review). One of the most intuitively plausible mediators of stereotype threat's effects on test performance is anxiety, or physiological arousal more generally. Although self-reports of anxiety level have failed to statistically mediate stereotype threat effects (Aronson et al., 1999; Oswald & Harvey, 2001; Schmader, 2002), direct physiological evidence suggests a role for arousal. In one study, for example, Black participants' blood pressure rose substantially in a stereotype threat condition, whereas the blood pressure of Black participants in a non-stereotype threat testing condition actually dropped from baseline (Blascovich et al., 2001). In other studies, decreased heart-rate variability (Croizet et al., 2004) and estimates of participants' anxiety (e.g., fidgeting) by independent raters blind to experimental condition (Bosson, Haymowitz, & Pinel, 2004) have statistically mediated stereotype threat effects.

Cognitive interference: Intrusive thoughts and thought suppression

Steele et al. (2002) suggested that stereotype threat may interfere with cognitive performance because of intrusive worries about fulfilling the group stereotype or being judged according to it; efforts to quell such intrusive thoughts may, in turn, simply exacerbate the distraction, according to research on thought suppression (Wegner, 1994). Empirical support for Steele et al.'s (2002) hypothesis exists. In one study, stereotype threat reduced participants' working memory capacity, as measured by a task in which women were asked to count vowels in sentences at the same time they memorized words and then recall as many target words as possible. Impaired performance on this measure of working memory capacity statistically mediated the women's subsequent performance decrements on a math test under stereotype threat conditions (Schmader & Johns, 2003). In another study, women in a math testing situation were asked, between each math problem, to write down whatever thoughts happened to come to mind. Participants in the stereotype threat condition reported a greater number of negative math-related thoughts than did participants in the control condition, and the quantity of these intrusive thoughts significantly predicted test performance impairment (Cadinu et al., 2005). Performance in the stereotype threat condition declined as the test went on, suggesting that a vicious cycle

between negative thinking and poor performance may occur under conditions of stereotype threat.

Beyond test performance decrements: Further damage from stereotype threat

Distancing oneself from a threatened identity

Even if the exact mechanisms of harm are not fully understood, stereotype threat clearly can hurt academic test performance. Recent research suggests that stereotype threat may also affect vulnerable individuals' identities; specifically, in response to the activation of negative stereotypes about a demographic group to which they belong, people may distance themselves from the threatened identity. In one study, for example, Black students under the threat of an academic ability test reported liking stereotypically African American items such as hip-hop and basketball less than did Black students not under stereotype threat or White students in either condition (Steele & Aronson, 1995). Over the long term, Black students who persevere in their academic efforts may feel pressured to abandon their Black social identity, while female students who focus on math may marginalize their feminine identity insofar as it fits the stereotype concerning women's mathematical abilities (see Pronin, Steele, & Ross, 2004). Thus, long-term stereotype threat sequelae may help to explain the greater levels of peer group ostracism that academically successful Black students experience compared to their high-achieving White peers (Fryer, 2006; Zirkel, 2004).

Disengagement from a threatening domain

In addition to the abandonment of valued social identities, stereotype threat may promote disengagement from threatening domains among stereotyped individuals (Steele, 1997; Steele et al., 2002). Not caring about academics is one way to avoid some of the pernicious effects of stereotype threat, but performance under such circumstances is likely to be even more depressed than it is under the load of stereotype threat when one cares about the domain and puts forth effort. Supporting the idea of long-term academic disengagement among Black students, Osborne (1997) found a weaker relationship between academic performance and global self-esteem among Black students than among White students. Black students, apparently, invest less of their self-worth in academic performance than do White students, perhaps due partly to stereotype threat (see also Major, Spencer, Schmader, Wolfe, & Crocker, 1998; Verkuyten & Brug, 2003).

Identifying moderators: What determines the strength of stereotype threat effects?

Test characteristics, real and purported

Researchers have identified several variables—some involving the test, some involving the test-taker, and some involving the testing environment—that moderate stereotype threat effects. For example, research suggests that more difficult tests produce greater stereotype threat effects (Steele et al., 2002). Given the crucial role that performance anxiety may play

in mediating stereotype threat, it is no surprise that frustratingly hard tests are most likely to induce stereotype threat. Independent of the actual properties of a test, stereotype threat effects can be exacerbated or attenuated by the representation of the test to examinees. As noted previously, stereotyped groups show greater performance decrements when a test is purported to show intergroup score differences or is represented as diagnostic of ability (e.g., Steele & Aronson, 1995).

Beliefs about intelligence

Striving in every testing situation to represent the test as nondiagnostic of ability may seem to be a daunting task. Fortunately, some research suggests that there may be ways to change students' chronic beliefs about the determinants of test performance and thereby alleviate stereotype threat over the long haul. In one study, Black college students who were persuaded to believe in the malleability of intelligence through effort performed better, academically, throughout the subsequent term than did their peers in the control group (Aronson, Fried, & Good, 2002). Similarly, seventh-grade girls who were taught that intelligence has more to do with one's efforts than with one's inherent, stable capacity did better on math tests compared to girls in the control condition—and, indeed, performed as well as the boys in their class (Good, Aronson, & Inzlicht, 2003).

Concerns about stereotypic evaluation by others

Stereotype threat may extend beyond self-doubt and involve concerns about other people judging oneself in terms of prevailing stereotypes. For this reason, Steele et al. (2002) argued that situations in which performance criteria are vague may promote stereotype threat; such situations suggest that the examiner possesses latitude to evaluate performance in terms of social stereotypes. A situation in which objective evaluation criteria are specifically delineated, on the other hand, may help to reduce stereotype threat. More generally, challenging students' beliefs that teachers or other evaluators view them through a stereotypic lens may be an effective means of alleviating stereotype threat. In one study, researchers were able to produce greater levels of trust, effort, and performance among Black students writing essays when the evaluators explicitly stated that they had extremely high standards but believed that the students could meet those standards (Cohen, Steele, & Ross, 1999). This kind of feedback—emphasizing high expectations—effectively contradicts the idea that one is in danger of being judged in terms of the negative intellectual stereotypes concerning one's demographic group. At the same time, this behavior may combat more general teacher expectancy effects that hurt stigmatized groups and help to maintain the Black–White score gap (see Ferguson, 2003; Kuklinski & Weinstein, 2001; McKown & Weinstin, 2002).

Social identity salience

A final moderator of stereotype threat strength is the salience of one's social identity. The more that a testing environment promotes awareness of one's threatened social identity, the greater the decrement in performance we can expect. In one study investigating this

idea, women completed math tests in classrooms that contained (along with numerous men) either zero, one, or two other female test-takers. Women performed worst when there were no other women present in the testing setting, whereas they performed best when two other women were present and there therefore was not as much attention called to their threatened identity as a woman in a mathematical setting (Inzlicht & Ben-Zeev, 2000).

This finding does not necessarily imply that testing environments should be segregated by sex, ethnicity, or other social categories in order to minimize social identity salience. A more ambitious long-term strategy might be reducing the threat posed by particular contexts for particular social identities—that is, increasing “identity safety” (Markus, Steele, & Steele, 2000; see also Cheryan, Plaut, Davies, & Steele, 2006) and reducing stigma. We discuss strategies for creating identity-safe classrooms later in this document.

Limitations of stereotype threat research

Population age

While acknowledging the reality of stereotype threat effects in lab studies using college samples, school psychologists may be skeptical about whether these effects generalize to natural school settings. One concern is population age. We know that children can discriminate gender by age 2 (Huston, 1987) and ethnicity by age 5 (Aboud, 1988; see also Quintana, 1998), and that they gradually acquire knowledge of cultural stereotypes from 6 to 10 (McKown & Weinstern, 2003; Ruble & Martin, 1998), but when exactly do they become vulnerable to the effects of stereotype threat? Knowledge of group stereotypes by itself may not be sufficient; a child may also have to understand the hierarchical nature of social categories (see Brown & Bigler, 2005).

In the first study examining stereotype threat in a school-age sample, Ambady, Shih, Kim, and Pittinsky (2001) found that girls as young as grades K–2 showed what the authors called “implicit stereotype awareness”—a tendency to assume that a top math student is male—even while they explicitly denied that girls and boys differed in math ability. These young girls and their counterparts in grades 6–8 subsequently succumbed to typical stereotype threat effects when gender was primed before test administration. Girls in grades 3–5, on the other hand, did not show implicit stereotype awareness and also did not show depressed performance under stereotype threat conditions, underscoring the need for more research examining this developmental period.

Whereas Ambady et al. (2001) directly activated children’s stereotypes by bringing attention to gender, McKown and Weinstein (2003) used an indirect form of stereotype induction, describing a test as either diagnostic or nondiagnostic of ability. In this study, too, the performance of young children (ages 6–10) who were aware of cultural stereotypes was impaired under stereotype threat conditions. Although more research using school-age samples is warranted, the existing evidence suggests that children are prone to the same stereotype threat effects as college students.

One unresolved question is whether high school students are as prone to stereotype threat as are younger children and college students. No research, to our knowledge, has addressed this issue, and it is possible that students who experience stereotype threat in their younger years—especially the types of students referred to school psychologists for

evaluation or intervention—may tend to disengage from the academic domain by high school, such that they are no longer vulnerable to stereotype threat. College students, of course, represent only those individuals who maintained some academic identification throughout their schooling, so the existence of stereotype threat in this population does not answer the question about high school students.

High-stakes versus laboratory testing

In real-world testing situations, where a single test score can determine whether one passes a class or gets into a desired college, students may be much more motivated to succeed than they are in lab settings where the chief motivation is to comply with the experimenter's instructions. It is unclear whether the higher stakes of real-world testing would magnify or attenuate the stereotype threat effects found in the social psychological laboratory. Unfortunately, this uncertainty creates an ethical quandary for researchers interested in investigating these questions. Introducing major stereotype threat manipulations into a high-stakes real-world test would mean potentially jeopardizing the academic futures of those vulnerable students who were randomly assigned to the stereotype threat condition. Nonetheless, Educational Testing Services researchers conducted one experiment that tested a stereotype threat prediction (Stricker & Ward, 2004). During an actual Advanced Placement (AP) Calculus examination, half of the students specified their ethnicity and gender before taking the test, while the other half specified their ethnicity and gender after completing the same test. The results did not support the predictions of stereotype threat theory; girls, for example, did not do worse when asked to indicate their gender before (rather than after) the test.

This study (Stricker & Ward, 2004), however, is a weak test of stereotype threat's relevance to the real world. Members of the negatively stereotyped groups in the study may have had the relevant stereotypes so activated already that the manipulation of indicating gender or ethnicity before taking the examination produced no additional change in stereotype activation. In other words, participants in both conditions may have been under substantial stereotype threat. This illustrates a more general problem with trying to demonstrate stereotype threat effects outside the laboratory. Tests such as AP exams and the SAT are so widely represented as diagnostic of ability—and gender and ethnic differences in score distributions so widely known—that the typical laboratory manipulations to reduce stereotype threat, such as labeling the test nondiagnostic, are unlikely to change students' perceptions of the tests (Cullen, Hardison, & Sackett, 2004; Steele et al., 2002).

Given the challenges of stereotype threat experiments in natural settings, Cullen et al. (2004) conducted a retrospective correlational analysis on real-world performance data. Regression lines representing prediction of criterion measures (e.g., class grades) by standardized tests did not show different slopes for different ethnic groups; the authors judged these results to be inconsistent with stereotype threat theory. Unfortunately, studies of this kind face a large hurdle: it is difficult to derive straightforward predictions about patterns of real-world performance from stereotype threat theory. Some authors (e.g., Cohen & Sherman, 2005) have suggested that stereotype threat may help to explain the well-known overprediction of college grades by SAT scores among Black students, but it is hard to understand why SAT performance should be less affected by stereotype threat than

college test performance, given that the SAT is difficult for the average student, is widely perceived to be a test of intellectual ability, and is very similar in content to the tests used in many laboratory-based stereotype threat studies. Indeed, given these facts, one might hypothesize from stereotype threat theory that SAT scores would *underpredict*—rather than overpredict—college grades for Black students. More thought needs to be given to the issue of how criterion and predictor variables might be differentially affected by stereotype threat if we wish for laboratory research to illuminate the causes of real-world performance discrepancies across the schooling and cognitive testing spans.

In summary, age-related inconsistencies and a paucity of studies in field settings warrant caution in the application of stereotype threat research to school psychological practice. Stereotype threat investigators themselves have emphasized (e.g., Steele & Aronson, 2004) that stereotype threat is not intended to provide a complete account of the test-score and achievement discrepancies currently observed between demographic groups such as Black and White students (e.g., Hyde, Fennema, & Lamon, 1990; Jencks & Phillips, 1998). To demarcate the precise empirical boundaries of stereotype threat—and to thereby elucidate its relevance to school psychological practice—future research must utilize well-validated measures in rigorously controlled studies of school-age children and adolescents in ecologically valid settings. In the meantime, despite the limitations of the extant stereotype threat literature, we believe that it does yield several tentative but important implications for school psychologists.

Implications for psychoeducational assessment

Based on stereotype threat research, we recommend five concrete changes to psychoeducational assessment with few foreseeable downsides. First, during the rapport-building period preceding test protocols (see Kamphaus, 1993; Sattler, 2001), psychologists must be careful not to inadvertently induce stereotype threat by asking questions about topics related to a student's demographic group (e.g., music preferences). Such questions may accidentally activate negative stereotypes about test performance or intellectual ability, even when the psychologist's assumptions about the student's group affiliations, knowledge, and interests are accurate. Second, and relatedly, measures of maximal performance (e.g., ability and achievement tests) should be placed at the beginning of assessments, before less formal self-report activities such as the clinical interview have a chance to activate stereotypes by covering topics such as family background, current home environment, preferred extracurricular activities, and self-perceptions of academic functioning. Third, psychologists should not describe the tests they administer as diagnostic of intellectual capacity or anything else, lest they promote test-takers' concerns about confirming group stereotypes of their abilities. Fourth, psychologists should consider directly measuring mediators of stereotype threat to determine whether it likely affected test performance. Measurement methods might range from informal interview questions to standardized measures of cognitive interference, state anxiety, and test anxiety (e.g., Spielberger, 1980). Finally, aided by direct measurements of stereotype threat mediators, psychologists must consider the possibility of stereotype threat when interpreting test scores of vulnerable group members. This directive is in concordance with APA and NASP guidelines for assessment, and, indeed, it helps to illustrate the critical distinction between

psychological *testing* and psychological *assessment* (cf. Matarazzo, 1990); in the latter, a trained psychologist interprets test scores or other information while considering the context of the individual client.

Implications for general education programming

Stereotype threat effects may extend into the classroom settings that yield grades and other information used by school psychologists to make decisions about students. Reforms of general educational programming—guided by research on the moderators and mechanisms of stereotype threat—may help to prevent stereotype threat from depressing the classroom test scores of stigmatized groups. School psychologists are well trained to take a leadership role in effecting changes in general education, and professional guidelines encourage participation in educational reform at a systems level (National Association of School Psychologists, 2000).

Knowledge of the stereotype threat literature and skills in applying this literature to classroom situations are especially timely in the context of recent high-stakes testing programs and educational accountability requirements now in effect for both general and special education students. As Thomas (2005) reported, mandates included in the No Child Left Behind Act and other legislation, such as requirements that school districts report achievement data by ethnic group, have brought group differences in test performance into bold relief and have increased general concern about reducing achievement gaps. Moreover, the criterion-referenced nature of these tests has led to “failure” rates that often vary across ethnic groups, prompting some community leaders to accuse school districts of bias and discrimination. We agree with Braden (2002) that school psychologists have much to contribute in this era of new educational accountability systems, and the stereotype threat literature may help lead to reductions in achievement gaps, a key goal of these systems.

Explaining standards

Drawing on the findings of Cohen et al. (1999) (see above), school psychologists might ask teachers to ensure that minority students (here we include female students) do not perceive poor grades to be due to their status as members of a group with a negative stigma concerning achievement. Teachers should always explain their grading criteria as explicitly as possible. Cohen et al.’s intervention—letting students know that challenging but attainable standards are present—is perhaps most crucial in the lower grades, before students have enough metacognitive awareness to judge their own competencies accurately. The substantial educational psychology literature on the importance of high expectations (e.g., Weinstein, 2002) makes the point more generally that students who think others expect them to do poorly are likely to perform in line with those perceived expectations.

With high-stakes standards-based assessments, especially, students should be given age-appropriate information about the nature of the assessment instruments and the reliable, objective methods with which they will be scored. Since some forms of these instruments ask for test-takers’ gender and ethnicity, students should be assured that the scorers typically will not have access to the examinees’ demographic characteristics. As a long-

term goal, psychologists might lobby for test developers to move demographic questions to the end of such examinations.

Teaching about stereotype threat

With older students, especially in schools with large minority populations, a frank discussion of stereotype threat may help to ameliorate possible stereotype threat effects. According to the research of Johns, Schmader, and Martens (2005), this educational intervention gives students a source to which anxiety during a test can be attributed, alleviating performance anxiety and partially preventing the depressed scores typical of minority students under stereotype threat. When school psychologists are asked to participate in or oversee high-stakes testing programs, in particular, they should consider developing presentations about stereotype threat, to be delivered either directly to the student body or through trained teachers.

Changing conceptions of intelligence

Intervention research demonstrating the moderation of stereotype threat effects by students' implicit theories of intelligence (Aronson et al., 2002; Good et al., 2003) suggests another classroom teaching topic: the malleability of intelligence and the importance of effort in producing academic achievement. Beyond its specific protective effects on students at risk for experiencing stereotype threat, viewing intelligence as changeable according to one's efforts is associated more generally with high academic achievement (e.g., Stipek & Gralinski, 1996), making Aronson et al.'s interventions potentially helpful for all students.

Identity-safe classrooms

The stereotype threat literature should lead teachers to create classrooms in which students do not feel defined or limited by their demographic group membership. The implications of this idea for multicultural education are complex and not always clear, since teachers' well-meaning efforts to respect students' cultural backgrounds can sometimes come across as ill-informed attempts to use students as "representatives" of their demographic groups. However, teachers *can* take proactive steps to show that students from all groups have the potential for academic excellence. Even brief interventions that call attention to members of diverse demographic groups who have succeeded in stereotypically "threatening" domains are sufficient to reduce performance anxiety and intrusive worries and, in turn, produce better test performance (McIntyre, Paulson, & Lord, 2003). Raising awareness of counterstereotypical exemplars of real-world achievement should especially boost the performance expectations of minority students in high-stakes assessment situations, where students may view the test as predictive of future life outcomes.

A second important part of creating identity-safe classrooms involves keeping students from having to choose between high academic achievement and peer acceptance. Conscious identification with one's ethnic group tends to be more central to the sense of self

in minority students than in White students (Tatum, 1997), making peer pressure against high achievement especially difficult for minority students to bear. Horvat and Lewis (2003) found that academically successful minority students employed a complex set of strategies to downplay their success around unsupportive peers, who might accuse them of “acting White,” while openly sharing their success with supportive peers. To allow students to perform at high levels without fear of reprisal by classmates, teachers and school psychologists must make active attempts to encourage a peer culture of high achievement by, for example, connecting academic success to powerful short- and long-term reinforcers valued by a wide variety of students.

Conclusion

Stereotype threat research suggests the ease with which powerful cultural expectations may be unintentionally conveyed during the course of a psychological assessment or a classroom examination. Ignoring the effects of cultural stereotypes on students’ academic functioning may compromise school psychologists’ competence to conduct nondiscriminatory evaluations and to increase schoolwide achievement levels. School psychologists and other educational officials therefore must attend to stereotype threat and its implications. The existing literature on stereotype threat is robust enough to allow for several recommendations to practitioners, but more research in real-world settings is needed to ascertain the external validity of stereotype threat theory. School psychologists can facilitate research in this area, and in doing so they may help their clients to achieve more at school and in life.

Acknowledgements

The authors thank Tanya Eckert for her comments on an earlier draft of this paper. This research was supported in part by a National Science Foundation Graduate Research Fellowship and a Regina Casper Stanford Graduate Fellowship awarded to the first author.

References

- Aboud, F. E. (1988). *Children and prejudice*. Oxford: Blackwell.
- Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science, 12*, 385–390.
- American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57*, 1060–1073.
- Aronson, J., Fried, C., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*, 113–125.
- Aronson, J., Lustina, M. J., Good, C., Koeogh, K., Steele, C. M., & Brown, J. (1999). When White men can’t do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*, 29–46.
- Ashton, M. C., & Esses, V. M. (1999). Stereotype accuracy: Estimating the academic performance of ethnic groups. *Personality and Social Psychology Bulletin, 25*, 225–236.
- Blascovich, J., Spencer, S. J., Quinn, D. M., & Steele, C. M. (2001). Stereotype threat and the cardiovascular reactivity of African Americans. *Psychological Science, 12*, 225–229.
- Bosson, J. K., Haymowitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology, 40*, 247–255.

- Braden, J. P. (2002). Best practices for school psychologists in educational accountability: High stakes testing and educational reform. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (pp. 301–319). (4th ed.). Bethesda, MD: National Association of School Psychologists.
- Brown, C. S., & Bigler, R. S. (2005). Children's perceptions of discrimination: A developmental model. *Child Development, 76*, 533–553.
- Brown, R. P., & Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology, 76*, 246–257.
- Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science, 17*, 572–578.
- Cheryan, S., Plaut, V. C., Davies, P., Steele, C. M. (2006). *The social representation of computer scientists: A barrier to inclusion for women*. Unpublished manuscript.
- Cohen, G. L., & Sherman, D. K. (2005). Stereotype threat and the social and scientific contexts of the race achievement gap. *American Psychologist, 60*, 270–271.
- Cohen, G. L., Steele, C. M., & Ross, L. D. (1999). The mentors' dilemma: Providing critical feedback across the racial divide. *Personality and Social Psychology Bulletin, 25*, 1302–1318.
- Croizet, J. C., & 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*, 588–594.
- Croizet, J. C., Despres, G., Gauzins, M. E., Huguet, P., Leyens, J. P., & Meot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin, 30*, 721–731.
- Cullen, M. J., Hardison, C. M., & Sackett, P. R. (2004). Using SAT–grade and ability–job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology, 89*, 220–230.
- Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28*, 1615–1628.
- Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology, 88*, 276–287.
- Ferguson, R. F. (2003). Teachers' perceptions and expectations and the Black–White test score gap. *Urban Education, 38*, 460–507.
- Frisby, C. (1999). Culture and test session behavior: Part II. *School Psychology Quarterly, 14*, 281–303.
- Fryer, R. G. (2006, Winter). "Acting white": The social price paid by the best and brightest minority students. *Education Next, 6*(1), 53–59.
- Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin, 28*, 659–670.
- Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Applied Developmental Psychology, 24*, 645–662.
- Horvat, E. M., & Lewis, K. S. (2003). Reassessing the "burden of 'acting white'": The importance of peer groups in managing academic success. *Sociology of Education, 76*, 265–280.
- Huston, A. C. (1987). The development of sex typing: Themes from recent research. In S. Chess & A. Thomas (Eds.), *Annual progress in child psychiatry and child development* (pp. 168–186). New York: Brunner/Mazel.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139–155.
- Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11*, 365–371.
- Jencks, C., & Phillips, M. (Eds.). (1998). *The Black–White test score gap*. Washington, DC: Brookings Institution.
- Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychological Science, 16*, 175–179.
- Kamphaus, R. W. (1993). *Clinical assessment of children's intelligence*. Boston, MA: Allyn & Bacon.
- Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. New York: Wiley.
- Kaufman, A. S. (2000). Tests of intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 445–476). New York: Cambridge University Press.

- Kuklinski, M. R., & Weinstein, R. S. (2001). Classroom and developmental differences in a path model of teacher expectancy effects. *Child Development*, *72*, 1554–1578.
- Lopez, E. C., & Rogers, M. R. (2001). Conceptualizing cross-cultural school psychology competencies. *School Psychology Quarterly*, *16*, 270–302.
- Major, B., & O'Brien, L. T. (2005). The social psychology of stigma. *Annual Review of Psychology*, *56*, 393–421.
- Major, B., Spencer, S. J., Schmader, T., Wolfe, C. T., & Crocker, J. (1998). Coping with negative stereotypes about intellectual performance: The role of psychological disengagement. *Personality and Social Psychology Bulletin*, *24*, 34–50.
- Markus, H. R., Steele, C. M., & Steele, D. M. (2000). Colorblindness as a barrier to inclusion: Assimilation and nonimmigrant minorities. *Daedalus*, *129*, 233–259.
- Mash, E. J., & Terdal, L. G. (1997). Assessment of child and family disturbance: A behavioral-systems approach. In E. J. Mash & L. G. Terdal (Eds.), *Assessment of childhood disorders* (pp. 3–70). New York: Guilford.
- Matarazzo, J. D. (1990). Psychological assessment vs. psychological testing: Validation from Binet to the school, clinic, and courtroom. *American Psychologist*, *45*, 999–1017.
- McIntyre, R. B., Paulson, R. M., & Lord, C. G. (2003). Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology*, *39*, 83–90.
- McKay, P. F., Doverspike, D., Bowen-Hilton, D., & Martin, Q. D. (2002). Stereotype threat effects on the Raven Advanced Progressive Matrices scores of African Americans. *Journal of Applied Social Psychology*, *32*, 767–787.
- McKown, C., & Weinstein, R. S. (2002). Modeling the role of child ethnicity and gender in children's differential response to teacher expectations. *Journal of Applied Social Psychology*, *32*, 159–184.
- McKown, C., & Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Development*, *74*, 498–515.
- National Association of School Psychologists. (2000). *Principles for professional ethics*. Bethesda, MD: Author.
- Ortiz, S. O. (2002). Best practices in nondiscriminatory assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV, Vol. 2* (pp. 1321–1336). Bethesda, MD: National Association of School Psychologists.
- Osborne, J. W. (1997). Race and academic disidentification. *Journal of Educational Psychology*, *89*, 728–735.
- Oswald, D. L., & Harvey, R. D. (2001). Hostile environments, stereotype threat, and math performance among undergraduate women. *Current Psychology: Developmental, Learning, Personality, Social*, *19*, 338–356.
- Padilla, A. M. (2001). Issues in culturally appropriate assessment. In L. A. Suzuki, J. G. Ponterotto, & P. J. Meller (Eds.), *Handbook of multicultural assessment* (pp. 5–27). (2nd ed.). San Francisco, CA: Jossey-Bass.
- Quintana, S. M. (1998). Development of children's understanding of ethnicity and race. *Applied and Preventive Psychology*, *7*, 27–45.
- Pronin, E., Steele, C. M., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: Women and mathematics. *Journal of Experimental Social Psychology*, *40*, 152–168.
- Ruble, D., & Martin, C. L. (1998). Gender development. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology, Vol. 3* (pp. 933–1016). New York: Wiley.
- Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego, CA: Author.
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, *38*, 194–201.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*, 440–452.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*, 4–28.
- Spielberger, C. D. (1980). *Test anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797–811.
- Steele, C. M., & Aronson, J. A. (2004). Stereotype threat does not live by Steele and Aronson (1995) alone. *American Psychologist*, *59*, 47–48.
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in experimental social psychology, Vol. 34* (pp. 379–440). New York: Academic Press.

- Stipek, D., & Gralinski, J. H. (1996). Children's beliefs about intelligence and school performance. *Journal of Educational Psychology*, 88, 397–407.
- Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology*, 77, 1213–1227.
- Stricker, L. J., & Ward, W. C. (2004). Stereotype threat, inquiring about test takers' ethnicity and gender, and standardized test performance. *Journal of Applied Social Psychology*, 34, 665–693.
- Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology*, 66, 21–36.
- Tatum, B. D. (1997). *“Why are all the black kids sitting together in the cafeteria?” and other conversations about race*. New York: Basic Books.
- Thomas, R. M. (2005). *High stakes testing: Coping with collateral damage*. Mahwah, NJ: Erlbaum.
- Valencia, R. R., & Suzuki, L. A. (2000). *Intelligence testing and minority students: Foundations, performance factors, and assessment issues*. Thousand Oaks, CA: Sage Publications.
- Verkuyten, M., & Brug, P. (2003). Educational performance and psychological disengagement among ethnic-minority and Dutch adolescents. *The Journal of Genetic Psychology*, 164, 189–200.
- Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, 101, 34–52.
- Weinstein, R. S. (2002). *Reaching higher: The power of expectations in schooling*. Cambridge, MA: Harvard University Press.
- Williams, R. L. (1971). Abuses and misuses in testing Black children. *The Counseling Psychologist*, 2, 62–73.
- Zenderland, L. (1998). *Measuring minds: Henry Herbert Goddard and the origins of American intelligence testing*. Cambridge: Cambridge University Press.
- Zirkel, S. (2004). What will you think of me? Racial integration, peer relationships and achievement among white students and students of color. *Journal of Social Issues*, 60, 57–74.