



The Gordon Commission
on the Future of Assessment in Education

A Social Psychological Perspective on the Achievement Gap in Standardized Test Performance Between White and Minority Students: Implications for Assessment

Rodolfo Mendoza-Denton
University of California, Berkeley

The National Study Group for the Development of Affirmative Academic Ability (Bennett, Bridglall, Cauce, Everson, Gordon, Lee, Mendoza-Denton, Renzulli, and Stewart, 2007) recently examined national trends in standardized tests for White, Black, and Hispanic students in the United States. The data selected for examination was broadly inclusive by design, and included both middle and high school samples, a range of years, and multiple academic subjects. More specifically, the group examined nationally representative Assessment of Education Progress (NAEP) scores for 8th and 12th grade students between the years 1996 and 2002 in science, math, and reading. The group also examined SAT scores for Black and White students between the years of 1996 and 2003.

For NAEP scores, the data revealed that across years, subjects, and grades — in other words, in *every single comparison* conducted — Black and Hispanic students scored significantly lower than their White counterparts (see [Figure 1](#)). For SAT scores, Black-White differences across eight years ranged from 92 to 108 points. The consistency of these achievement differences is as striking as it is sobering, and made even more so by the fact that they replicate decades of data demonstrating achievement gaps in standardized tests among these groups (e.g., The College Board, 2010). These achievement gaps are the central focus of this chapter.

Scholars have debated the source and meaning of these academic achievement gaps intensely. While psychologists tend to interpret these achievement gaps as examples of test bias and unfairness due to the fact that one should not, a priori, have reason to expect factors such as race to influence test scores (e.g., Helms, 2006, 2009; Walton and Spencer, 2009), the testing field rejects the existence of such differences as evidence of bias per se (Sackett, Borneman, and Connelly, 2008, 2009). Sackett et al. (2008) put forward the proposition that, in fact, such differences may reflect construct-relevant variance: ethnic group membership, for example, may be associated with undervaluing academic achievement and/or spending less time studying. In the words of Sackett et al. (2008), “It is precisely because of these potential alternative explanations that the dominant view in the testing field rejects the position that a finding of a relationship between race ... and test scores can be directly interpreted as signaling bias or unfairness” (p. 223). This position reflects the strong assumption that if a student is unmotivated or spends less time studying than his/her peers, this reflects a lack of personal responsibility and, as such, is fairly reflected on the standardized test. Sackett et al. (2009) expand on this notion of

personal responsibility, arguing “we find it hard to imagine any instructor responding positively to a student who says, ‘I chose not to read the assigned material covered on the test, so it’s unfair that you penalize me for my incorrect responses’” (p. 286).

I reviewed research suggesting that, even when it came to processes that reflect personal choice — such as time spent studying or becoming invested in academics — the interpersonal milieu surrounding the student has a profound influence on the development of these processes. The aim is to shed light on how certain classes of students, in particular ethnic minority students, are systematically exposed to toxic messages about their abilities that force many to minimize the importance of academics (Major, Spencer, Schmader, Wolfe, and Crocker, 1998; Ogbu, 1991; Steele, 1997). This, in turn, suggests that academic achievement gaps, rather than reflecting test bias, may be more accurately described as reflecting societal bias. I conclude with some possible directions that fair testing may take in the context of a cultural context that does not support all students to the same degree.

This chapter explores the issue of academic motivation and performance from a social psychological perspective. I focus specifically on the academic achievement gap between White and minority students, not only because it represents one of the widest and most persistent gaps in testing, but also because social-psychological research in this area has, to date, concentrated most heavily on this dimension. Nevertheless, I discuss gaps related to other social identities (e.g., gender, SES) where relevant research sheds light on the present narrative. The hope is that educators and stakeholders within the testing domain will find the insights reviewed here helpful toward progress across a range of social identities.

Academic Engagement and Self-Protection

A critical part of my argument lies in the notion that toxic environmental messages about a student’s (or a group’s) abilities can force students to adopt self-protective strategies that involve privileging certain identities over others in the self-concept. Foundational research in anthropology by John Ogbu and colleagues (Ogbu and Simons, 1998; Ogbu, 2008) provides theoretical and empirical background for this idea. Ogbu’s (1978) cross-cultural comparative work strongly contradicts potential arguments of dispositional differences in cognitive capacity as an explanation for academic achievement gaps (e.g., Herrnstein and Murray, 1994), finding that while stigmatized minorities in one country (e.g., the Boraku and Koreans in Japanese

schools) show academic underperformance, these same groups show no such underperformance in contexts where these identities are not stigmatized (e.g., in U.S. schools). Ogbu proposes a socio-cultural explanation for achievement differences, arguing instead that among stigmatized minorities a set of culturally shared protective narratives emerges. These narratives recognize the unequal resources, treatment, and expectations of a majority-dominated educational system toward minority learners, and the lack of trust and inclusion that these practices engender. Ogbu and Simons (1998) note that an instrumental response to these inequities is a shift in attitudes about schooling, thereby questioning the value of academic rites of passage and credentials and, by extension, the symbols of that system, such as standardized tests. The socio-cultural nature of these adaptations is further underscored by the fact that recently immigrated minorities (who do not yet share this worldview) and voluntary minorities (members of groups who immigrated voluntarily to the land where they now are a minority) do not exhibit as dire academic underperformance (Ogbu and Simons, 1998).

In Ogbu's writing, we recognize a symbiosis of socio-cultural beliefs (e.g., stereotypes) that happen "outside of the head" and intrapersonal beliefs (e.g., about the value of schooling) that happen "inside the head." While social constructivist and cultural psychological approaches (e.g., Gergen, 1990; Mendoza-Denton and Ayduk, in press) have long recognized that individual psychology is inseparable from social context, this view has stood in contrast to a traditional view that separates "person" forces from "environmental" forces (Mendoza-Denton and Mischel, 2007; Ross and Nisbett, 1991). The notion that a student's declaration that "I chose not to read the assigned material" is a matter of personal responsibility implicitly works within the latter concept, but viewing it within the former framework suggests that the educational environment can play a role in its genesis and in its change.

While John Ogbu's groundbreaking work has been sharply criticized as an exercise in victim blame (M.A. Ogbu, 2008), his theory and research frame an oppositional stance against majority-dominated schooling as a socio-cultural — and here I argue psychological — adaptation to the injustices that involuntary minorities, in particular, have historically faced when intersecting with a White educational system. The casting of Ogbu's research as blaming the victim may be partly due to his explicit acknowledgment that outside forces such as discrimination have an effect on a target's own psychological processes. However, even though Ogbu recognized that disenfranchisement from schooling involves choice and personal beliefs, it

is critical to recognize that Ogbu did not see this choice as the “fault” of students. Rather, he saw socio-structural change, as psychologists and educators tend to see today, as the route to change.

Psychological Adaptation: The Regulation of Self-Esteem and the Self-Concept

In the section that follows we pursue a particularly personal dimension of self that is both intertwined with the social environment and has played a central role in the scientific narrative around the effects of prejudice on its targets. This dimension is self-esteem.

Historically, self-esteem has been a central topic for psychologists interested in understanding prejudice and its effect on targets. “There is ample evidence of inferiority feelings and morbid self-hate in all minority groups,” wrote Erik Erikson (1956, p. 115), and Cartwright (1950) observed that “self-hatred and feelings of helplessness tend to arise from membership in underprivileged or outcast groups.” Consistent with this historical zeitgeist, Herb and Mimi Clark (1947) conducted a study that showed young Black children preferred to play with Black dolls over White dolls, and that this preference was related to children’s perceptions of the Black dolls as dirtier, uglier, and less desirable than the White dolls.

The United States Supreme Court specifically cited Clark and Clark (1947) in its 1954 landmark *Brown vs. Board of Education* ruling that declared segregation in school settings unconstitutional. Nevertheless, although the abolishment of school segregation helped the nation move forward, the implicit notion that stigmatized minorities are passive recipients of the prejudiced and stereotypical messages that they continuously face remains very much entrenched in the popular imagination (APA, in press). Clark and Clark’s (1947) study remains cited today as evidence that stigmatized group members internalize the negative stereotypes that are directed at them — that is, that they come to believe the stereotypes and their self-concept suffers as a result (Davis, 2005).

Modern research demonstrates, however, that the relationship between self-esteem and stigmatization recognizes that people, while sensitive to negative information about themselves, do not necessarily allow negative messages about them or their groups to affect them. Instead, they are active participants in the formation and regulation of their self-concept. Leary (2000), for example, reviews extensive empirical evidence for a view of self-esteem as a “sociometer” — a type of social barometer that rises when one feels accepted and dips when one is rejected. In a related vein, Tajfel and Turner’s (1986) influential Social Identity Theory proposes that people

derive self-esteem from one's group memberships, such that the accomplishments and social standing of the groups one is psychologically affiliated with (e.g., a sports team, an organization) are reflected in one's self-esteem. Together, these interpersonal notions of self-esteem suggest that people are active participants in the management of their self-esteem. Research finds, for example, that people derogate out-groups (i.e., groups that they don't subjectively belong to) and glorify in-groups to maximize their self-esteem: in more phenomenological language, the belief that my group is better than your group helps me feel good about myself. In addition, if a certain possible self is problematic, either because it is viewed in negative terms by society or because one's membership in the group is questioned, people can choose to disown that particular identity — a process known as disidentification. To anticipate an idea that is developed more fully in a subsequent section: when a student chooses to disidentify from the academic enterprise, this choice reflects a sense that one is not fully accepted as a member of the academic enterprise in the first place. This has direct implications for those involved in the development and management of standardized tests, who often are seen as the gate-keepers for participation in the various rites of the academic enterprise (e.g., who goes to what college or professional school).

In spite of the widespread belief that minorities have lower self-esteem than majority group members, self-esteem data for African Americans do not support this assumption. A recent extensive meta-analysis (Twenge and Crocker, 2002), in fact, shows that African Americans today actually have higher levels of self-esteem relative to European Americans. How does one make sense of these data? Crocker and Major (1989) proposed a process whereby stigmatization can protect self-esteem. Although this may seem counter-intuitive at first glance, it makes sense when one takes into account the idea that people actively manage and protect their self-concept. Crocker and Major proposed that when a person from a stigmatized minority group receives negative feedback — for example, a bad grade on a homework assignment or a low mark on a test — and the stigmatized identity is salient, it precipitates a state of attributional ambiguity — that is, of not knowing whether to attribute the negative outcome to one's own shortcomings or to bias on the part of the evaluator. The ambiguity results from the fact that each of these explanations is plausible — and self-esteem protection is achieved by attributing negative outcomes to prejudice.

Crocker, Voelkl, Testa, and Major (1991) tested this idea empirically. In one study, for

example, the researchers had Black and White participants take part in a purported study on friendship development. When participants arrived to the lab, they were told that another participant was already sitting in an adjacent room on the other side of a one-way mirror that could see into the participant's own room. Half of the participants were in the "blinds down" condition: they were told that the blinds would come down so that physical appearance did not interfere with friendship development. In the "blinds up" condition, participants were told that the blinds would stay up because appearance is an important and natural part of friendship development.

With this setup, the participants were further divided into two groups, such that half of them received a profile from the other participant that indicated they really wanted to meet them (positive feedback) or indicated they were not at all eager to meet them (negative feedback). Self-esteem was measured before and after the manipulations, allowing the researchers to examine changes in self-esteem.

The results for the Black participants in the study are illustrated graphically in [Figure 2](#). As the figure illustrates, in the "blinds down" condition, the findings correspond well with the sociometer view of self-esteem — when the other person was eager to meet them, participants' self-esteem rose, whereas when the other person rejected them, their self-esteem dropped from baseline levels. The pattern in the "blinds up" condition, however, supported the protective properties of stigma hypothesis: when participants received negative feedback their self-esteem remained intact, as if discounting the views of their partner given that this person was aware of their race. It is important to note that the researchers were able to make this argument because no such discounting occurred in the "blinds down" condition. Major, Kaiser, and McCoy (2003) similarly found that attributions to sexism buffered both men and women against depression in the face of a negative evaluation. The other striking finding from this study, which was originally not pursued in research on attributional ambiguity, was that when given positive feedback in the "blinds up" condition, the self-esteem of Black participants actually fell. This also indicates a different type of discounting that is just as damaging — the discounting of positive feedback as politically correct but insincere (see Piff and Mendoza-Denton, in press).

Boundary conditions. It is important to note that research has uncovered important boundary conditions for the protective properties of stigma phenomenon. These boundary conditions are important as we consider the role that prejudice, stigma, and self-esteem play in

relation to academic achievement. The first of these boundary conditions is people's subjective sense of being responsible for their stigmatized condition. More specifically, when people feel that they are somehow responsible, or to blame for the prejudice directed against them, self-esteem is unlikely to be protected in the face of ambiguous prejudice. Crocker, Cornwell, and Major (1993), for example, documented a loss of self-esteem following social rejection among overweight participants, even when they made an attribution that the rejection was due to weight. Similar findings have been shown for the stigma of mental illness, where the perception that one is somehow responsible for one's condition is commonly endorsed (Hinshaw, 2007; Martinez and Mendoza-Denton, in press). The second boundary condition is that, even in the face of a stigma that is unarguably outside of one's responsibility or control, people do not discount the prejudice (and thus protect self-esteem) in the absence of a broader social narrative supporting this discounting. Twenge and Crocker (2002), for example, show that the self-esteem of African Americans relative to European Americans rose considerably after the Civil Rights movement, which made the unfairness of racial prejudice chronically accessible as an explanation for the negative outcomes experienced by many African Americans (see [Figure 3](#)). Twenge and Crocker did not observe patterns of self-esteem protection among African Americans who grew up prior to the Civil Rights movement, or among Asian Americans more generally, who have not had the benefit of a widely accepted, collective consciousness-raising movement in U.S. history (Chan and Mendoza-Denton, 2008).

Linking Attributional Ambiguity to Academic Engagement

How can research on attributional ambiguity help us understand the academic achievement gap in general, and patterns of academic engagement more specifically? Attributional ambiguity plays a large role in today's educational settings (London, Downey, Romero-Canyas, Rattan, and Tyson, 2012). More specifically, while overt, hostile discrimination and explicitly discriminatory laws and practices are increasingly rare, today's environment that is intolerant toward discrimination, coupled with the continued negative attitudes and stereotypes against minorities that prevail to this day have created a unique situation where racial discrimination has gone "underground" (e.g., Dovidio and Gaernter, 2000). Stated another way, explicit affirmations of egalitarianism in laws, educational policy, and personal beliefs today stand in stark contrast to other markers of discrimination that include discrimination on a

personal (e.g., through microaggressions, Sue, 2010) as well as an institutional level (e.g., numerical underrepresentation of minorities in many settings). These subtle, and — importantly — ambiguous signals of unequal treatment, as much as overt discrimination, can trigger attributional ambiguity which, in turn, may lead to mistrust of educators and the educational enterprise. Consistent with this claim, Whitt, Nora, Edison, Terenzini, and Pascarella (1999) have shown that perceptions of an unwelcoming educational climate are related to students' decreased understanding of educational material, while Litzler, Lange, and Brainard (2005) have similarly linked such perceptions to lower educational commitment.

Mendoza-Denton, Goldman-Flythe, Pietrzak, Downey, and Aceves (2010) have recently provided experimental evidence of these processes specifically as they relate to educational outcomes. We invited African American students to the lab to purportedly participate in a study of student-instructor relations and academic feedback. We asked the students to write a position essay on a controversial topic, and led them to believe that an actual instructor at the university (whom the students thought was White) was in another room waiting to evaluate their essays. The students were randomly assigned to receive either positive or negative feedback, and were also randomly assigned to fill out a demographic form in which they either did or did not disclose their race (crossing these two factors yielded four experimental cells in the study). The results confirmed that students who were high on race-based rejection sensitivity (Mendoza-Denton, Downey, Purdie, Davis, and Pietrzak, 2002) — that is, concerned about being targets of discrimination — and who disclosed their race in the demographic form seemed impervious to the feedback they received from the instructor, as measured by changes in their self-esteem. By contrast, when they did *not* reveal their race, all of the students' academic self-esteem rose in response to positive feedback, and fell in response to negative feedback. These results are compelling for two reasons — first, they *directly* link concerns about discrimination to engagement or disengagement from academic enterprises such as writing essays and getting feedback on them. Secondly, the findings show that even students who are concerned about discrimination, when given an environment in which they can trust that their identity will not be used against them, can attach their self-concept to academic outcomes (a point that often is lost in discussions that broadly characterize African American students as endorsing an oppositional culture).

To review: people are not simply passive recipients of negative attitudes about the groups with which they are affiliated. Targets of discrimination think about the causes, meanings, and sources of the negativity directed against them, and decide individually and collectively on self-protective courses of action. As we have reviewed, personal disengagement from domains in which feedback is not trusted can protect self-esteem. Importantly, however, it also involves a trade-off: to the degree that feedback is valid and informative, important opportunities for growth and development are lost in the discounting of feedback (Piff and Mendoza-Denton, in press; Steele, 1997). The trade-off that attributional ambiguity represents, where learning opportunities are forced to compete in a zero-sum context against potentially biased feedback, is an example of societal bias because this trade-off is not experienced by majority group members (APA, in press). This points to a strong need to create educational environments in which people can trust the fairness of the feedback they receive, as well as their own belongingness within these environments (Mendoza-Denton et al., 2010; Cheryan, Plaut, Davies, and Steele, 2009; Walton and Cohen, 2007; 2011).

Social Identity Threat

As I argued in the last section, research does not support the internalization of stigma hypothesis for African Americans; rather, the data are more consistent with the self-protective properties of stigma view. This, however, raises a new question: does this data not suggest that ethnic minorities should be free to continue academic pursuits without concerning themselves about prejudice?

Research conclusively shows that the answer to this last question is “no.” While research on attributional ambiguity shows that attributions to discrimination can protect self-esteem, these attributions do not protect against intense emotions such as anxiety, anger, and the feeling that one is not in control of one’s own outcomes (Major et al., 2003; Levy, Freitas, Mendoza-Denton, Kugelmass, and Rosenthal, 2010). These emotions not only can affect health outcomes (as evidenced by racial disparities in cardiovascular illness; Blascovich, Spencer, Quinn, and Steele, 2001), but also can affect academic outcomes. In this section, we review a psychological phenomenon that helps explain the link between stereotypes, negative emotions, and academic achievement. This phenomenon is known as Stereotype Threat, and more recently as Social Identity Threat.

In one of the earliest demonstrations of this phenomenon, Steele and Aronson (1995) presented both White and Black students with questions such as those that they might find on a standardized achievement test. It is important to note, at the outset, that all students in the study completed the exact same test questions. The students' race constituted one of the factors in the study; the second factor was an experimental framing manipulation. In one condition (the "ability diagnostic" condition), the students were told that the researchers were interested in verbal ability and that the test contained items that were diagnostic of this ability. In the other condition (the "non-diagnostic" condition), the task was framed differently: the students were told that questions would help researchers understand how people solve problems, and that the researchers were not interested in evaluating the participants' ability. This latter manipulation, by design, was meant to relieve the students of the psychological threat that as members of a negatively stereotyped group, their ability was under suspicion or scrutiny. Again, all students then proceeded to answer the very same questions, albeit differently framed, and their performance was examined while controlling statistically for their prior SAT scores (this was done to allow the findings to reflect the effect of the manipulation, independently of prior skills or preparation).

The findings revealed that the White students in the study performed comparably regardless of which condition they were in. The African American students, however, showed a remarkable sensitivity to the diagnosticity message. Specifically, they performed just as well as the White students in the non-diagnostic condition, but underperformed relative to White students in the "ability diagnostic" condition. For the African American students, a small, but psychologically critical, framing of the test questions was enough to affect performance, such that when the questions were framed as diagnostic of one's intelligence, the achievement patterns mirrored the group differences that constitute the academic achievement gap. These findings are illustrated in [Figure 4](#).

Spencer, Steele, and Quinn (1999) replicated these effects with gender. Their experimental manipulation involved telling men and women in the diagnostic condition that the test they were about to take revealed gender differences, thus directly bringing to bear a specific social identity (gender) to the test. In the other condition, the men and women were told that the questions revealed no gender difference. Replicating the findings from Steele and Aronson (1995), Spencer et al. (1999) found that the gender differences led women to underperform

academically, whereas the scores of men and women were statistically indistinguishable in the gender-neutral condition.

This underperformance effect was labeled stereotype threat because the psychological threat associated with confirming a stereotype, or the aura of suspicion around one's abilities, is enough to cause performance decrements. It is an aversive state, where anxiety and intrusive ideation about the stereotype essentially “get in the way” of students' being able to focus on the questions (Schmader, Johns, and Forbes, 2008; Steele, 1997). Importantly, as emphasized by Steele (1997), one does not need to believe the stereotype to be affected by it — one only has to be threatened by the possibility that one might be judged by others in light of that stereotype. By removing this threat, the performance differences disappear. These data flat out contradict notions of genetic or dispositional differences to explain test performance gaps because biological differences should be reflected in spite of psychological framing.

Stereotype threat effects have been replicated for a range of stigmatized social identities, including socioeconomic status (Croizet and Claire, 1998), mental health status (Quinn, Kahng, and Crocker, 2004), and even caste status in India (Hoff and Pandey, 2004). Beyond the value of replicating the phenomenon empirically, these studies suggest that it is not necessarily just concerns about confirming a stereotype that have an effect on students, but more broadly, a feeling that a self-relevant social identity is under threat. As such, Steele, Spencer, and Aronson (2002) coined the more general term “social identity threat” to refer to this “state of psychological discomfort that people experience when confronted by an unflattering group or individual reputation in situations where that reputation can be confirmed by one's behavior (Aronson and McGlone, 2009, p.154).”

Coping with Social Identity Threat

Much like coping with attributional ambiguity, coping with stereotype and social identity threat involves trade-offs. Studies have shown that those students who are most vulnerable to stereotype threat are precisely those students who are most identified within the domain. Steele (1997) for example, reviewed research showing that stereotype threat effects were most pronounced among academically identified students, with non-identified students seemingly protected against the effects of threat. Major and colleagues (1998) as well as Steele (1997) have shed light on the psychological roots underlying these data, and their conclusions echo

arguments that the reader should recognize. These scientists distinguish between situation-specific psychological disengagement from a threatening, aversive situation involving identity threat, for example a standardized testing situation (Major et al., 1998). With repeated exposure to such threat, however, situational disengagement morphs into domain disidentification — that is, removing one's sense of identity more broadly from the evaluative domain and not allowing success or failure in that domain to affect one's self-concept (Steele, 1997). Reminiscent of Ogbu's arguments, then, disidentification can be seen as a form of coping that can protect from stereotype threat. But the trade off is clear: disidentification can close doors to achievement in that domain. A choice to disidentify from a domain in which one's identity is stigmatized — STEM (Science, Technology, Engineering, and Math) fields, for example, if one is a woman — becomes more like a forced choice in the interests of one's mental health.

Such choices have a self-perpetuating, status-quo prolonging quality to them. The underrepresentation of women in STEM fields remains a serious societal issue (Cheryan et al., 2009; London et al, 2012), with the number of women earning Ph.D.'s in STEM fields lagging dramatically behind the number of women earning bachelors' degrees in these fields (NSF, 2009). As argued earlier, these inequities are both a consequence and a reflection of societal stereotypes that women are bad at math ("Math is hard!" chirped a Barbie Doll that was briefly on the market in 1992) and importantly, communicate hegemonic, exclusionary cultural norms and practices. For example, Cheryan et al. (2009) found that relatively innocuous cues that tend to be associated with men in computer science (e.g., a Star Trek poster) were enough to activate belonging concerns among women and reduced their interest in the field. Plaut, Thomas, and Goren (2009) found greater minority employee satisfaction in organizations that upheld a cultural norm of multiculturalism over a norm of colorblindness. Inzlicht and Ben-Zeev (2000) found a strong relationship between the ratio of men to women present in the setting where women had to complete a stereotype-relevant task, and the performance of women in this task: the fewer other women in the room, the worse women performed.

Of note is that many of the cues above (e.g., Star Trek posters, a distinct lack of diversity in a given setting, a colorblind norm) are ambiguous with respect to discrimination, and often co-exist alongside explicit regulations or endorsements of the value of diversity. These types of contradictory signals are common in many of today's majority-dominated educational and occupational institutions. Research suggests that such ambiguity can exert a negative influence

on performance because it leads to a process of rumination and uncertainty about one's perceptions. Mendoza-Denton, Shaw-Taylor, Chen, and Chang (2009), for example, had female test takers complete a competency exam in the purported office of a male interviewer who was late for their interview. The décor of the room was manipulated so as to communicate the attitudes of the interviewer. In one room (the egalitarian room), the décor consisted of decorations that suggested the interviewer valued gender equality, and included a poster for breast cancer awareness as well as pictures of his daughter. In the chauvinist room, the decorations suggested the interviewer held sexist attitudes, and included pictures of bikini-clad models on motorcycles and cases of Big Daddy IPA beer. Finally, in a third room, the decorations were ambiguous, revealing the interviewer's male identity and his position of power but not betraying anything concrete about his gender attitudes. Consistent with the idea that ambiguity is disruptive, the results showed that among women who were high on gender-based rejection sensitivity (London et al, 2012), or chronically concerned about being rejected on the basis of their gender, performance on the competency exam was most disrupted in the ambiguous room — even more so than in the explicitly chauvinist room. It is also important to note that even though gender-based rejection sensitivity is a dispositional, individual differences variable, the point of the above analysis is not to indicate who is versus who is not concerned about discrimination (and thus, potentially bringing charges of “why can't you just get over it?”), but rather, to illustrate the processes through which toxic environments and their perception thereof can disrupt performance.

Test Bias and Social Identity Threat

Although research on social identity threat has provided extensive independent evidence for the robustness of this phenomenon, as reviewed above, the utility of such processes for understanding group level achievement differences, including the standardized test achievement gap, outside of the laboratory setting has been questioned. As discussed earlier, this debate is centrally concerned as to whether group-level achievement differences (e.g., as a function of gender, race, ethnicity, socioeconomic status) in high-stakes “cognitively loaded tests of knowledge” can be interpreted as indicators that the tests are biased (Sackett et al., 2008; Walton and Spencer, 2009).

The central argument against test bias is that evidence for such bias should be seen in the

under-prediction of achievement from these tests for affected populations. In other words, test bias would be seen in the depression of test scores relative to one's actual ability, and thus the same test score should predict higher achievement for the student that the test is biased against. An example helps concretize the idea: if a 1,200 on an SAT predicts a grade point average of 3.5 in college for a White student, for example, the same 1,200 SAT score should predict a higher GPA for a Black student. In their analysis of large scale national data, Sackett and colleagues found evidence for a slight under-prediction of scores for women, but clear evidence of over-prediction for minority students. In other words, for minority students, the same level of standardized test performance relative to White students was associated with lower outcome criterion performance, namely grade point average (GPA).

Although the argument that over-prediction is evidence against test bias (and, by extension, does not challenge the validity of the tests as indicators of core competence in a domain), this argument assumes that the outcome indexes that the tests are compared to are free of the effects of bias. In other words, the argument makes the strong assumption that the educational environment in which students earn their GPA, in the example above, is the same for all students.

If this were indeed the case, the logic of under-prediction would hold. However, as we have seen, social psychological research strongly suggests that the interpersonal educational environment is psychologically not equivalent for minority and majority students. This is particularly true in majority dominated educational settings (Bowen and Bok, 1998). The historical legacy of discrimination and exclusion based on group membership affects people's sense of belonging in the educational environment, and in the larger educational enterprise. As such, this interpersonal dimension is extremely difficult to account for among developers and reformers of standardized tests, precisely because of the explicit interest in standardization across settings. We review the interpersonal dimension in the section below before returning to the issue of test bias.

The Interpersonal Dimension: Sensitivity, Valuation, and Belonging

"Gradually I began to think of myself as a social psychologist. With this change in self-concept came a new accountability; my self-esteem was affected now by what I did as a social psychologist, something that hadn't been true before... [An] observer might say that even though my background was working-class, I had special advantages: achievement-oriented parents, a

small and attentive college. But these facts alone would miss the importance of the identification process I had experienced: the change in self-definition and in the activities on which I based my self-esteem. They would also miss a simple condition necessary for me to make this identification: treatment as a valued person with good prospects.”

From *Race and The Schooling of Black Americans* (Steele, 1992)

An emergent theme that ties together the literature summarized thus far, from oppositional identity to attributional ambiguity to social identity threat, is that discrimination engenders a sense of mistrust among its targets that they will be treated with fairness and respect, and viewed as valued members of a common group. Concerns about interpersonal valuation and trust are related to — but can affect targets independently of — concerns about confirming negative stereotypes or beliefs. For example, following the 1996 passage of Proposition 209 banning the consideration of race, ethnicity, or sex in California university admissions, its effects continue to reverberate across the University of California educational system. In 2006, for example, out of a freshman class of more than 4,400 students at UCLA, only 100 were African American. In 2011, in anticipation of the reconsideration of Proposition 209, a group of students organized a “diversity bake sale,” in which the same baked goods were priced differently for members of different racial/ethnic groups. Their idea was to protest race-conscious admissions decisions and it sparked widespread protests across the university (see Mendoza-Denton, 2011). The students who protested the offensive bake sale gained admission to UC Berkeley under proposition 209, and had every reason to feel comfortable that they earned their way into Berkeley through their own achievements. Nevertheless, two themes in their protest were their continued invisibility (“Don’t UC us?” read one sign) and marginalization (see [Figure 5](#)). Clearly, issues of belonging continue to be central to the experience of these students even when they have no reason to worry about confirming negative stereotypes about inferior ability.

To help account for the impact of such concerns along the interpersonal dimension, Mendoza-Denton and colleagues (Mendoza-Denton et al., 2002; Mendoza-Denton, Pietrzak, and Downey, 2008) proposed that the psychological legacy of discrimination, exclusion, or mistreatment leads to a heightened awareness and anxiety, at the individual level, that one may be subjected to similar prejudice in future occasions. This dynamic — status-based rejection sensitivity — already has been mentioned in this chapter, and it has been applied to the stigmatized social identities of African Americans (Mendoza-Denton et al., 2002), Latino/as

(Mendoza-Denton and Page-Gould, 2008; Page-Gould, Mendoza-Denton, and Tropp, 2008), Asian Americans (Chan and Mendoza-Denton, 2008), women (London et al., 2012), gay men (Pachankis, Goldfried and Ramrattan, 2008), and the elderly (Chow, Au, and Chiu, 2009). It belongs to the family of psychological dynamics relating anticipated discrimination to anxiety and threat. However, it emphasizes the relational origins of these concerns, as the dynamic is modeled after theories of attachment and close interpersonal relationships (see Downey and Feldman, 1996; Mendoza-Denton and Ayduk, in press).

The status-based rejection sensitivity model explicitly equates the experience of discrimination as akin to the experience of social rejection, with a strong precedent for this parallelism in the literature (Branscombe, Schmitt, and Harvey, 1999). The model proposes that, as a result of vicarious or direct experiences of discrimination, people develop anxious expectations that they will be similarly rejected in the future. These anxious expectations lead to a heightened sensitivity to environmental cues relating to discrimination, and prime the perceptual system to perceive this noxious stimulus more readily (echoing the processes related to ambiguity and attribution described above). Perceived discrimination leads to a cascade of intense, negative cognitive and affective reactions, including anger, rumination, and physiological stress (Page-Gould et al., 2008). The anticipation of potential discrimination engenders a sense of mistrust and a lack of belonging, similar to the way that possible rejection engenders insecurity and relationship trouble in romantic relationships (Mendoza-Denton and Ayduk, in press). In the educational context, race-based rejection expectations have been directly linked to an inability to fully focus on the educational mission that students arrived at the institution to pursue in the first place.

Mendoza-Denton and colleagues (2002) provided empirical evidence for this process among African American students in a predominantly White educational setting. Students high on race-based rejection sensitivity (RS-race), during the first 21 days of their college experience, showed slight but significant differences in the ways that they intersected with the university and its representatives. Whereas African American students low in RS-race showed increased liking and trust for their professors over the first three weeks of college — a pattern to be expected as a function of familiarity — students higher in RS-race did not evidence these increases. Students higher in RS-race also reported feeling less inclusion and identification in the university, a pattern that was magnified at the end of their first, and subsequent, years of college (Mendoza-

Denton et al., 2002). Moreover, the slight but significant difference in liking toward professors translated into a very tangible subsequent outcome; namely, reduced attendance at review sessions, increased anxiety interacting with professors and TAs, and reduced GPA (see [Figure 6](#)). Aronson and Inzlicht (2004), in an independent analysis, linked RS-race to an unstable academic self-concept.

An important aspect of the RS-race perspective is the trajectory of students lower in RS-race, who do not evidence these patterns of intertwining acceptance and rejection concerns at the educational institution with academic achievement. Indeed, students lower in RS-race, who were not as worried about discrimination, show significantly better adjustment and academic achievement. As stated earlier, the focus is not on who is or may be high versus low in RS-race (a person-centric analysis), but rather, on the fact that the data link the absence of race-based rejection concerns to increased academic achievement. Thus, the onus is on the fostering and maintenance of educational environments in which all students can feel that their race (or other stigmatized identity) is valued within the educational context (Mendoza-Denton et al., 2009; APA in press).

Walton and Cohen (2007, 2011) also have demonstrated the impact of belonging concerns for academic achievement. Incoming minority students at a predominantly White institution with a prominent exclusionary history were exposed to a brief belonging intervention. Specifically, the students in the treatment condition were exposed to an intervention designed to mitigate their doubts about belonging in the university. They were presented with information in their first year at the school that students from all backgrounds and ethnicities experienced anxiety over their fit and their belonging at school, and that these concerns dissipated over time. The intervention was delivered in a brief single session but, as research shows, small differences in behavior at critical transition junctures snowball over time to have long term impact (Ruble and Seidman, 1996). In the case of Walton and Cohen, their brief belonging intervention resulted in increased academic achievement three years after its delivery, as well as greater perceived health and well-being (Walton and Cohen, 2011, see [Figure 7](#)).

A similar theme about the importance of the interpersonal environment comes from a different intervention conducted by Mendoza-Denton and Page-Gould (2008). Here, the treatment condition consisted of a friendship intervention, in which friendship was experimentally induced between randomly assigned Latinos/as and White same-sex pairs.

Compared to a same-race friendship intervention, the cross-race friendship intervention increased the sense of belonging and satisfaction within the university among students high in RS-race. The data suggest that in the presence of potentially discriminating cues in the environment (as indexed, here, by RS-race), a close interpersonal connection with a representative of the dominant, majority group decreases anxiety about interactions with other members of the group (Page-Gould et al., 2008). This, in turn, opens the door for more positive interactions and intellectual pursuits within the institutional setting.

An important detail of a previously discussed study is now timely. Recall that Mendoza-Denton and colleagues (2010) conducted a study in which African American participants received either positive or negative feedback about their essays from a purported White instructor at the university, and either disclosed or did not disclose their own race to this purported evaluator (see page 13). The findings showed that in the presence of race-based rejection concerns, self-esteem was impervious to feedback, whereas in the presence of identity safety, students allowed their self-esteem to become contingent on the feedback of the evaluator. While the findings were discussed earlier in terms of their relevance for self-esteem, the relevant detail for the current discussion is that the findings further showed that the effects of the manipulations on self-esteem were mediated, or explained, through the students' sense that the feedback they received communicated that they were valued, respected, and accepted. This is exactly the sentiment captured so beautifully in the quote by Claude Steele (1992) that began this section.

Revisiting the Issue of Test Bias

We now turn our attention back to the issue of test bias, over-prediction, and under-prediction. The argument made in the last section is that relational concerns surrounding belonging, acceptance, and valuation in relation to one's race pervade majority-dominated educational environments, as well as the symbolic tokens of those environments (such as, I would argue, standardized tests).

But how can one, in real-world settings, compare outcome and criterion fairly when bias pervades every level of the educational enterprise? Walton and Spencer (2009) attempted to tackle this question directly by adopting a strategy in which they compared the performance of students in contexts pervaded with threat versus contexts free of such threat relative to standard

criteria. Using meta-analytic techniques, Walton and Spencer compared threatening versus non-threatening environments in both experimental and real-world intervention contexts, improving on the correlational techniques on which Sackett and colleagues (2004, 2008) relied. The results showed that standardized tests do, in fact, under-predict the performance of minority students by about a fifth of a standard deviation. As the authors note, “The observed effect sizes suggest that the SAT Math test underestimates the math ability of women like those in the present sample by 19 to 21 points, and that the SAT Math and SAT Reading tests underestimate the intellectual ability of African and Hispanic Americans like those in the present sample by a total of 39 to 41 points for each group (p. 1137).” The results suggest that the underestimation of performance among stigmatized groups does not result from test bias per se, as we have anticipated throughout this chapter, as much as from the threatening contexts in which such tests are administered and taken. This is of considerable import for stakeholders in standardized testing reform.

What About Intelligence?

A topic that we have not yet directly discussed, but pervades the entire conversation around group difference and achievement, is the role of intelligence — how we construe it, how we measure it, and our beliefs about it. In a recent interview, for example, James Watson, who was awarded the Nobel Prize for his work on the helical structure of the DNA molecule, noted that he was “inherently gloomy about the prospects of Africa,” explaining that “all our social policies are based on the fact that their intelligence is the same as ours — whereas all the testing says not really (Nugent, 2007).” Watson’s comments are a cold reminder that many in society, including prominent thinkers, scientists, and policy-makers, regard tests of intellectual ability as faithful markers of ability, as diagnostic of an immutable quality as a blood test is for blood type. In this section, I review evidence for the position that this framing of intelligence and competency testing intersects with negative stereotypes about the intellectual inferiority of ethnic minority students in ways that further discourage participation and engagement among minority students in standardized testing.

Research in psychology suggests that the belief in intelligence as a fixed, dispositional entity, despite being widely endorsed in modern society (Martinez and Mendoza-Denton, in press) is just that — a belief. This is not to say that it does not have important consequences

(quite the opposite, as beliefs are tightly linked to action), but it does suggest that it is socially constructed, upheld, and perpetuated by the culture and its institutions (Dar-Nimrod and Heine, 2006; Mendoza-Denton and Mischel, 2007). It also means that as a belief, it is amenable to revision — and in such revision we can find reason to hope.

Research by Carol Dweck and colleagues (see Dweck and Leggett, 1988; see Dweck, 2002; Molden and Dweck, 2006) has demonstrated that in spite of the normative view, there are individual differences in the ways that people construe intelligence. More specifically, some uphold the view that intelligence is a fixed, immutable entity, and as such have been called entity theorists. By contrast, others uphold the view that intelligence is malleable, and that one can incrementally increase intelligence through hard work. This group has been dubbed incremental theorists.

Entity theorists, who believe that intelligence is fixed, approach learning and education with different strategies. Early work (Elliot and Dweck, 1988) showed that children who are entity theorists are more likely to interpret their performance on tests as diagnostic of their intelligence, and as such are highly invested in performance (e.g., getting a good grade). These performance goals are adopted with the end goal of proving that one is, in fact, intelligent. The logically consistent corollary of this mindset is that if one gets a bad grade or performs badly on a test, it can only mean that one must not be smart. Unsurprisingly, then, entity theorists tend to shy away from learning opportunities that involve difficulty for fear of making mistakes and being labeled as unintelligent.

By contrast, within the mindset of incremental theorists, who believe you can grow your intelligence, the meaning of difficult learning tasks is diametrically different: it is an opportunity for growth. This differential approach toward challenge is particularly important in schooling, since schooling almost by definition means being exposed to material that one has not yet mastered. Illustrating this idea, Grant and Dweck (2003) examined the performance of students in a highly competitive college chemistry course (such courses, it should be noted, are often gateway courses used to funnel or weed out students for more advanced classes). The students who endorsed learning (as opposed to performance) goals not only found the material more intrinsically motivating, but also were protected from the inevitable setbacks of the enterprise: if they did poorly on one test, students with learning goals, but not those with performance goals, dramatically improved their performance on a subsequent exam.

Blackwell, Trzesniewski, and Dweck (2007) reported similar findings, but strengthened the causal argument that intelligence beliefs cause performance differences through an experimental intervention among 7th grade students. The treatment group received lessons and demonstrations about the malleability of intelligence, with vivid demonstrations of neurons making connections during the process of learning. The control group taught students about the structure of memory with equally engaging material. Although both groups' grades showed evidence of decline prior to the intervention (7th grade being a time when many students' grades decline), the intervention group reversed this trend, whereas the control group continued this downward trajectory (see Figure 8).

When tasks are easy or one has not yet received feedback, having an entity theory can be a powerful motivator, as it allows people to feel smart. As noted, however, when faced with the setbacks that learning inevitably entails, an entity theory can become a risk factor. This point is nicely illustrated by Mueller and Dweck (1998), who either praised children for their intelligence or for their hard work following the children's success in an easy task. Following this praise manipulation, all children were exposed to a difficult set of problems that nobody could solve — a setback manipulation. Following this setback manipulation, the children were once again given a similarly easy task as in the first phase of the study. The results were telling: even though all children were praised after doing well on the first task, the children who confronted failure after having been praised for their intelligence were less likely to want to persist on and enjoy the new problems, and performed worse compared to the children who were originally praised for their hard work. These results strikingly demonstrate how messages of fixed intelligence (“You sure are smart!” goes a song in the popular children's show *Blue's Clues*), may, in fact, provide the opposite outcome to what many parents and educators surely intend when uttering praise.

The Intersection of Intelligence Beliefs and Negative Ability Stereotypes

A central component of the negative stereotypes of ethnic minorities in this country is the suspicion of low intellectual capacity. Thus, much of the research and outcomes related to entity theories of intelligence become immediately relevant to the discussion of stereotypes and achievement as well. Aronson, Fried, and Good (2002) explicitly linked the literatures on stereotype threat and theories of intelligence, noting that minority students who are under threat of confirming a stereotype adopt many of the same defensive strategies as entity theorists do,

including choosing easier tasks, focusing on performance rather than learning, and disengaging from difficulty (Major et al., 1998).

Accordingly, Aronson and colleagues conducted an intervention aimed at altering minority students' theories of intelligence. African American college students were exposed to an intervention similar to the one described above by Blackwell and colleagues (2007), emphasizing that intelligence is “like a muscle” that grows and is engaged in the face of difficulty and challenge. Replicating and reaffirming the findings from research on theories of intelligence, Aronson et al. (2002) found that students exposed to the incremental intervention not only showed greater academic enjoyment and engagement relative to students in a control condition, but also achieved higher grade point averages. As Aronson et al. conclude, “negative ability stereotypes may derive part of their power to undermine intellectual performance and motivation precisely because they imply a self-threatening and inalterable deficiency — a fixed lack of intelligence (p. 116).”

Mendoza-Denton, Kahn, and Chan (2008) expand this idea with evidence that assumptions of fixed intelligence may contribute to a widening of the academic achievement gap. Specifically, while a belief in fixed intelligence hampers performance in the context of a negative stereotype, the researchers reasoned that this same belief should bolster performance in the presence of a positive stereotype. By believing that one's advantage is fixed (e.g., men are good at math), stereotypes of fixed advantage may facilitate performance by easing performance concerns and ensuring that one's advantage cannot be taken away. To test this idea, Mendoza-Denton et al. (2008) experimentally manipulated a) whether a positive stereotype of math ability was confirmed or disconfirmed; and b) whether math ability was framed as fixed or malleable. The researchers tested the effects of these manipulations separately in two groups of students who are favorably stereotyped as being good at math — men (relative to women) and Asians (relative to Whites). The results confirmed the hypotheses: math performance was highest when the stereotype of high ability was framed as being both true and immutable ([see Figure 9](#)). As the authors conclude, “To the degree that the educational system reaffirms an entity view of intellectual abilities through ability tracking and intelligence testing ... the current findings suggest an exacerbation and maintenance of performance gaps between groups about whom stereotypes exist. Further implications of the current study may be felt in areas such as career choice, such that entity-minded individuals may overselect favorably stereotyped domains in

which performance is boosted. A cycle in which a favorably stereotyped groups' success then confirms societal expectations, and perpetuates inequities, understandably follows (p. 1192)."

Implications and New Directions for Assessment

In this section, we delve into the implications of the above research for assessment in educational contexts and, particularly, educational assessments. Admittedly, the lessons from this collective literature are difficult to incorporate into assessment practices: the locus of the issue, as we have observed, is more precisely located in the context and the environment (i.e., threatening context, societal bias) rather than in the test (i.e., test bias). Nevertheless, I argue that the testing field should not conclude "this is not my problem," since the achievement gaps in testing, and their implications for diversity and opportunity, both contribute to and perpetuate toxic environments.

The recommendations in this section fall into two general categories: 1) changes in test development procedures; and 2) structural changes and interventions that testing services and organizations can promote and fund. These recommendations are not mutually exclusive.

Category I: Changes in Test Development Procedures, with an example

An over-emphasis on terms such as "aptitude," "cognition," or "ability" are likely to automatically activate stereotype threat among minority students (Steele, 1997), and tests should be renamed with a nod to the extensive research on this phenomenon. Beyond the names given to the tests, however, testing organizations should explore the potential for new, unexplored indexes of effectiveness or performance, and in particular, indexes that do not show a priori differences in performance as a function of group membership. This strategy, in a way, is like working backward from the outcomes that do not show bias and predict effectiveness to the test items that index these outcomes. This is a powerful logic and, fortunately, one that has evidence-based, empirical data in its favor.

Recently, Shultz and Zedeck (2008, 2011) completed a long-term research study to create not only a more effective test of lawyering effectiveness than the current standard in the field (the LSAT), but also a more equitable one for minorities and women. As Shultz and Zedeck note, LSAT scores are heavily relied upon by law schools in their admissions decisions and, consequently, this test functions as a gatekeeper into the profession. The LSAT predicts student

performance well, particularly in the first year of law school, because the LSAT assesses the skill set (e.g., memorization) that law schools emphasize early on in training. However, as the researchers found, the skill set required for practicing lawyers, as determined both by lawyers and by professors, is much broader than the skill set measured by the LSAT. As such, it's not a surprise that the LSAT does a much poorer job predicting lawyer performance outside of law school (Shultz and Zedeck, 2011).

Through extensive field research that included interviews and questionnaires, Shultz and Zedeck arrived at a set of 26 skills that law school alumni, clients, faculty, students, and judges deemed as important for lawyering. These include factors like practical judgment, negotiation skills, the ability to see the world through the eyes of others, developing relationships within the legal profession, and strategic planning ([see Table 1](#)). Consistent with the finding that the LSAT predicts lawyer effectiveness modestly, LSAT scores predicted only eight of the 26 factors — with two being negatively related to LSAT scores.

A second phase involving item development and identification led to measures that could predict these 26 criteria. For example, Shultz and Zedeck developed Situational Judgment Tests (SJT) in which test takers have to decide the best way to handle a set of critical situations related to the law profession (e.g., “You learn that a co-worker, Angela, who you helped train for the job, copied some confidential and proprietary information from the company’s files. What would you do?”). Lievens and Sackett (2012) have recently similarly advocated for the use of situational judgment tests as a valid strategy to broaden testing criteria. Shultz and Zedeck also identified a set of biographical/experiential factors (e.g., “How many times in the past year were you able to think of a way of doing something that most others would not have thought of?”), as well as dispositional factors (e.g., ambition, interpersonal sensitivity, and dispositional optimism) that predicted the previously identified 26 skills of lawyer effectiveness. In the end of this laborious but worthwhile process, the researchers had a validated, concrete alternative to the LSAT that a number of law schools are now implementing into their admissions procedures (Shultz and Zedeck, 2011).

In sum, rather than focusing on the possibility of test bias in available assessments, it is possible not only in principle, but also in concrete practice, to seek indicators of performance in a given domain that do not show evidence of group differences. This idea paves the way for a new generation of assessment materials that include a wider variety of skill sets, a trend that is

gaining traction in the field of testing.

Category II: structural changes

Dissemination of incremental intelligence views. As the research highlighted in this review suggests, promoting the cultural adoption of the belief that intelligence is malleable is an important way to promote achievement and to reduce academic achievement gaps among groups for whom a suspicion of fixed ability is applicable. This is a particularly attractive angle for wide dissemination because it does not target, blame, or focus on specific group. Dissemination at this level may be targeted at students, as reviewed here (Aronson et al., 2002; Blackwell et al., 2007), which has the advantage of empowering students and potentially providing them with coping tools that they can take with them across contexts. Such interventions also can be aimed at educators, who are likely to treat students differently if they believe the students can learn (e.g., Rosenthal and Jacobsen, 1968). A third prong in the dissemination approach involves media campaigns with the direct aim of getting the word out that intelligence is not a fixed entity. In the age of Facebook, Youtube, and the World Wide Web, campaigns that specifically advertise to at-risk groups is a potentially valuable tool. Testing organizations may have the resources necessary to fund the professional development of such tools, and to jump-start their dissemination.

Explicit advocacy by testing organizations against tracking practices. Ability tracking (Hallinan, 1994) consists of separating students into different learning tracks in elementary, middle, and high schools based on measures of intellectual ability. As Martinez and Mendoza-Denton (in press) note, tracking in practice often leads to lifelong, divergent educational trajectories, in which students in initially lower tracks end up with fewer opportunities for advancement than do their high-track counterparts. As such, tracking may foster self-fulfilling prophecies, as exemplified in the research reviewed here, where students labeled as being low in ability essentially go on to fulfill this promise (Merton, 1957). This is especially problematic when minority group members are overrepresented in low ability groups. Research on the success of the jigsaw classroom (e.g., E. Aronson, 2002), in which students at different levels of knowledge or ability more successfully learn by being the group experts for different parts of the course material, and by switching flexibly in their roles as learners and teachers, shows that learning is not incompatible with a diversity of knowledge or skill level in the classroom.

Increasing representation across all levels of testing organizations. As the research

described in this review shows, seemingly innocuous cues such as the numerical representation of minority group members in a testing situation (Inzlicht and Ben-Zeev, 2000), or the décor in a particular testing space (Cheryan et al., 2011; Mendoza-Denton et al., 2010) can affect the performance of stigmatized group members. While the focus has been on underrepresentation and negative representations of minority group members in educational settings, the converse idea — that increasing representation and having positive visible figures in such settings — should boost performance. From the perspective of testing organizations, the hiring and promotion of a diverse workforce takes on magnified symbolic as well as practical importance.

Fine, Weis, and Powell (1997) make an important distinction between numerical diversity, however, and relational diversity, which recommends tolerance and celebration of different points of view and emphasizes relational bonds across group memberships (Mendoza-Denton et al., 2008). A related concept that has received attention in the psychological literature recently is that of multiculturalism, an intergroup ideology that recognizes differences and upholds them as a specific strength of the collective. As mentioned earlier, Plaut, Thomas, and Goren (2009) find that minority employees in organizations that endorse multiculturalism as a company value feel more trust and commitment toward the organization than do their peers in more colorblind environments, where differences are devalued in favor of a superordinate — often mainstream — identity. Research shows that diversity in work contexts is related to increased creativity and better problem solving (see Bowen and Bok, 1998), suggesting that efforts within testing organizations to increase relational diversity also should benefit the central aim of such organizations in terms of the products that are developed.

The recommendations in category II may be met with skepticism as not being part of the job of testing organizations. While understandable, it is also the responsibility of testing organizations to provide assessment tools that accurately capture differences in competence and qualifications. As such, the field of testing also may take an interest in promoting the conditions that most accurately reflect such differences. It is my sincere hope that this review has been helpful to spark ideas and considerations of how the field of assessment may more fruitfully approach the persistent academic achievement gap.

References

- American Psychological Association, APA Presidential Task Force on Reducing and Preventing Discrimination Against and Enhancing Benefits of Inclusion of People Whose Social Identities Are Marginalized in U.S. Society. (2012). *Dual Pathways to a Better America: Preventing Discrimination and Promoting Diversity*. Washington, DC: Author.
- Aronson, E. (2002). Building empathy, compassion, and achievement in the jigsaw classroom. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pgs. 209 – 225). San Diego, CA: Academic Press.
- Aronson, J., Fried, C.B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, 38, 113–125.
- Aronson, J., & Inzlicht, M. (2004). The ups and downs of attributional ambiguity: stereotype vulnerability and the academic self-knowledge of African American college students. *Psychological Science*, 829–836.
- Aronson, J., & McGlone, M. (2009). Stereotype and social identity threat. In T. Nelson (Ed.), *The handbook of prejudice, stereotyping, and discrimination*. New York: Guilford.
- Bennett, A., Bridglall, B., Cauce, A. M., Everson, H., Gordon, E. W., Lee, C. D., Mendoza-Denton, R., Renzulli, & Stewart (2007). All students reaching the top: strategies for closing academic achievement gaps. In E. W. Gordon & B. Bridglall (Eds.), *Affirmative development: Cultivating academic ability* (pgs. 239–275). Lanham, MD: Rowman & Littlefield.
- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, 78, 246-263.
- Blascovich, J., Spencer, S.J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science*, 12, 225-229.
- Bowen, D. and Bok, G. (1998). *The shape of the river: Long term consequences of considering race in college and university admissions*, Princeton: Princeton University Press.
- Branscombe, N.R., Schmitt, M.T., & Harvey, R.D. (1999). Perceiving pervasive discrimination among African Americans: Implications for group identification and well-being. *Journal of Personality and Social Psychology*, 77, 135-149.
- Cartwright D. (1950). Emotional dimensions of group life. In M.L. Raymert (Ed.), *Feelings and Emotions* (pgs. 439–447). New York: McGraw- Hill
- Chan, W., & Mendoza-Denton, R. (2008). Status-based rejection sensitivity among Asian Americans: implications for psychological distress. *Journal of Personality*, 76, 1317-1346.

Chow, D. Sau-king, & Au, E. W. M. (2008). Predicting the psychological health of older adults: Interaction of age-based rejection sensitivity and discriminative facility. *Journal of Research in Personality*, 42, 169-182. doi: 10.1016/j.jrp.2007.05.001.

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology*, 97, 1045-1060. doi:10.1037/a0016239

Chow, D. Sau-king, & Au, E. W. M. (2008). Predicting the psychological health of older adults : Interaction of age-based rejection sensitivity and discriminative facility. *Journal of Research in Personality*, 42, 169-182. doi: 10.1016/j.jrp.2007.05.001.

Clark, K.B., & Clark, M (1947). Racial identification and preference in Negro children. In T. M. Newcomb and E.L. Hartley (Eds.), *Readings in Social Psychology* (pgs. 169-178). New York: Holt.

College Board. (2010). *2010 college-bound seniors: Total group profile report*. Available from <http://professionals.collegeboard.com/profdownload/2010-total-group-profile-report-cbs.pdf>

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review*, 96, 608-630.

Crocker, J., Voelkl, K., Testa, M. & Major, B. (1991). Social stigma: The affective consequences of attributional ambiguity. *Journal of Personality and Social Psychology*, 60, 218-228.

Crocker, J., Cornwell, B., & Major, B. (1993). The stigma of overweight: Affective consequences of attributional ambiguity. *Journal of Personality and Social Psychology*, 64, 60–70.

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.

Dar-Nimrod, I., & Heine, S. J. (2006). Exposure to scientific theories affects women's math performance. *Science*, 314, 435.

Davis, K. (2005). *A Girl Like Me*. Reel Works Teen Filmmaking.

Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism in selection decisions: 1989 and 1999. *Psychological Science*, 11, 315-319. doi:10.1111/1467-9280.00262

Downey, G., & Feldman, S. (1996). Implications of rejection sensitivity for intimate relationships. *Journal of Personality and Social Psychology*, 70, 1327-1343.

Erikson, E. (1956). The problem of ego-identity. *Journal of the American Psychoanalytic Association*, 4, pgs. 56–121.

- Dweck, C. S. (2002). The development of ability conceptions. In A. Wigfield & J. Eccles (Eds.), *The development of achievement motivation*. New York: Academic Press.
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256-273.
- Elliott, E., & Dweck, C. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, 54, 5-12.
- Fine, M., Weis, L. & Powell, L.C. (1997). Communities of difference: a critical look at desegregated spaces created for and by youth. *Harvard Educational Review*, 67, 247-284.
- Gergen, K. J. (1990). Social understanding and the inscription of self. In Stigler, J.W., Shweder, R.A., & Herdt, G. (Eds.), *Cultural Psychology: Essays on Comparative Human Development* (pgs.569-607). Cambridge: Cambridge University Press.
- Grant, H., & Dweck, C. S. (2003). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology*, 85, 541–553.
- Helms, J. E. (2006). Fairness is not validity or cultural bias in racial-group assessment: A quantitative perspective. *American Psychologist*, 61, 845–859.
- Helms, J. E. (2009). Defense of tests prevents objective consideration of validity and fairness. *American Psychologist*, 64, 283–284.
- Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.
- Hinshaw, S. (2007). *The mark of shame: Stigma of mental illness and an agenda for change*. New York: Oxford University Press.
- Hoff, K. & Pandey, P. (2004) Belief systems and durable inequalities: an experimental investigation of Indian caste. *Policy Research Working Paper; no. 3351*, The World Bank.
- 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), 365–71. doi:10.1111/1467-9280.00272.
- Leary, M. (2000). *The nature and function of self-esteem: Sociometer theory*. Burlington: Elsevier.
- Levy, S.R., Freitas, A.L., Mendoza-Denton, R., Kugelmass, H., & Rosenthal, L. (2010). When sociopolitical events strike cultural beliefs: Divergent impact of Hurricane Katrina on African Americans' and European Americans' endorsement of the Protestant Work Ethic. *Basic and Applied Social Psychology*, 32, 207-216. DOI: 10.1080/01973533.2010.495673.

Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology*, 97, 460–468. doi:10.1037/a0025741

Litzler, E., Lange, S., and Brainard, S. (2005). Climate for Graduate Students in Science and Engineering Departments. *Proceedings of the 2005 American Society for Engineering Education Annual Conference and Exposition*.

London, B., Downey, G., Romero-Canyas, R., Rattan, A., & Tyson, D. (2012) Gender-Based Rejection Sensitivity and Academic Self-Silencing in Women. *Journal of Personality and Social Psychology*.

Major, B., Kaiser, C. R., & McCoy, S. K. (2003). It's not my fault: When and why attributions to discrimination protect self-esteem. *Personality and Social Psychology Bulletin*, 29, 772-781.

Major, B., Spencer, S., Schmader, T., Wolfe, C. & Crocker, J. (1998). Coping with negative stereotypes about intellectual performance: The role of psychological disengagement. *Personality & Social Psychology Bulletin*, 24, 34-50.

Martinez, A.G., & Mendoza-Denton, R. (in press). The Prospect of Plasticity: Malleability Views of Group Differences and their Implications for Intellectual Achievement, Mental/Behavioral Health, and Public Policy. *Social Issues and Policy Review*.

Mendoza-Denton, R. (2011, September 26). *Making the Invisible Visible: The Berkeley Campus Republicans' Bake Sale*. Psychology Today Online, Retrieved from <http://www.psychologytoday.com/blog/are-we-born-racist/201109/making-the-invisible-visible-the-berkeley-campus-republicans-bake-sal>

Mendoza-Denton, R., & Ayduk, O. (in press). Personality and social interaction: Interpenetrating processes. In K. Deaux and M. Snyder (Eds). *Oxford Handbook of Personality and Social Psychology*. New York: Oxford University Press.

Mendoza-Denton, R., Downey, G., Purdie, V. J., Davis, A., & Pietrzak, J. (2002). Sensitivity to Status-Based Rejection : Implications for African American Students' College Experience Origins of Anxious Rejection Expectations. *Journal of Personality and Social Psychology*, 83(4), 896 -918. doi: 10.1037//0022-3514.83.4.896.

Mendoza-Denton, R., Goldman-Flythe, M., Pietrzak, J., Downey, G., & Aceves, M. J. (2010). Group-Value Ambiguity: Understanding the Students' Self-Esteem. *Science*. doi:10.1177/1948550609357796.

Mendoza-Denton, R., Kahn, K., & Chan, W. Y. (2008). Can fixed views of ability boost performance in the context of favorable stereotypes? *Journal of Experimental Social Psychology*, 44, 1187–1193.

- Mendoza-Denton, R., & Mischel, W. (2007). Integrating system approaches to culture and personality: The cultural cognitive-affective processing system (C-CAPS). In S. Kitayama & D. Cohen (Eds), *Handbook of Cultural Psychology* (pgs. 175-195). New York: Guilford.
- Mendoza-Denton, R., & Page-Gould, E. (2008). Can cross-group friendships influence minority students' well-being at historically white universities? *Psychological Science*, 19(9), 933-9. doi: 10.1111/j.1467-9280.2008.02179.x.
- Mendoza-Denton, R., Pietrzak, J., & Downey, G. (2008). Distinguishing institutional identification from academic goal pursuit: interactive effects of ethnic identification and race-based rejection sensitivity. *Journal of personality and social psychology*, 95(2), 338-351. doi: 10.1037/0022-3514.95.2.338.
- Mendoza-Denton, R., Shaw-Taylor, L., Chen, S., & Chang, E. (2009). Ironic effects of explicit gender prejudice on women's test performance. *Journal of Experimental Social Psychology*, 45(1), 275-278. Elsevier. doi: 10.1016/j.jesp.2008.08.017.
- Merton, R.K. (1957). *Social theory and social structure*. New York: The Free Press.
- National Science Foundation, Division of Science Resources Statistics (2009). *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2009*, NSF 09-305, (Arlington, VA; January 2009). Available from <http://www.nsf.gov/statistics/wmpd/>.
- Molden, D.C., Dweck, C.S. (2006). Finding 'meaning' in psychology: a lay theories approach to self-regulation, social perception, and social development. *American Psychologist*, 61, 192-203.
- Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology*, 75, 33-52.
- Nugent, H. (2007, October 17). Black people 'less intelligent' scientist claims. *The Times Online*. Retrieved from <http://www.timesonline.co.uk/tol/news/uk/article2677098.ece>
- Ogbu, J. (1978). *Minority education and caste: the American system in cross-cultural perspective*. New York: Academic Press.
- Ogbu, J. (1986). The consequences of the American caste system. In U. Nessler (Ed.), *The school achievement of minority children: New perspectives*. Hillsdale, NJ: Erlbaum.
- Ogbu, J. (1991). Minority coping responses and school experience. *The Journal of Psychohistory*, 18, 433-456.
- Ogbu, J., & Simons, H. (1998). Voluntary and involuntary minorities: A cultural-ecological theory of school performance with some implications for education. *Anthropology and Education Quarterly*, 29, 155-188.
- Ogbu, J. (2008). *Minority status, oppositional culture, and schooling*. New York: Taylor and Francis.

Ogbu, M.A. (2008). *A note from Marcellina Ada Ogbu*. In J. Ogbu, *Minority status, oppositional culture, and schooling* (p. xxvii-xxx). New York: Taylor and Francis.

Quinn, D., Kahng, S., & Crocker, J. (2004). Discreditable: Stigma effects of revealing mental illness history on test performance. *Personality and Social Psychology Bulletin*, 30, 803 – 815.

Pachankis, J., Goldfried, M., & Ramrattan, M. (2008). Extension of the rejection sensitivity construct to the interpersonal functioning of gay men. *Journal of Consulting and Clinical Psychology*, 76, 306–317. doi:10.1037/0022-006X.76.2.306

Page-Gould, E., Mendoza-Denton, R., & Tropp, L. (2008). With a little help from my cross-group friends: Reducing intergroup anxiety through cross-group friendship. *Journal of Personality and Social Psychology*, 95, 1080-1094.

Piff, P.K., & Mendoza-Denton, R. (in press). Mixed signals: Culture and construal in the provision of feedback across group boundaries. In R. Sutton, M. Hornsey, and K.M. Douglas (Eds.), *Feedback: The communication of praise, criticism, and advice*. New York, NY: Peter Lang.

Plaut V.C., Thomas K.M. & Goren M.J. (2009). Is multiculturalism or color blindness better for minorities? *Psychological Science*, 20, 444-446.

Rosenthal, R, and Jacobsen, L. (1968). *Pygmalion in the classroom: teacher expectation and pupils' intellectual development*. New York: Holt, Rinehart and Winston.

Ross, L., Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. New York, NY, England: McGraw-Hill Book Company.

Ruble, D. N., & Seidman, E. (1996). Social Transitions: Windows into social psychological processes. In E.T. Higgins & A.W. Kruglanski (Eds.), *Social Psychology: Handbook of Basic Principles*, (pgs. 830-856). New York, NY: Guilford Press.

Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High-stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist*, 63, 215–227.

Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2009). Responses to issues raised about validity, bias, and fairness in high-stakes testing. *American Psychologist*, 64, 285–287.

- Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2004). On interpreting stereotype threat as accounting for African American-White differences on cognitive tests. *American Psychologist*, 59, 7–13.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115, 336–356.
- Shultz, M.M., & Zedeck, S. (2008). Identification, Development, and Validation of Predictors for Successful Lawyering. Unpublished manuscript, University of California, Berkeley.
- Shultz, M.M., & Zedeck, S. (2011). Predicting lawyer effectiveness: broadening the basis for law school admission decisions. *Law & Social Inquiry*, 36, 620–661.
- 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.
- Steele, C.M. (April 1992). Race and the schooling of black Americans. *Atlantic Monthly*, 68–78.
- 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, 613–629.
- Steele, C.M. & Aronson, J. (1995). Stereotype threat and the intellectual performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797–811.
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pgs. 379–440). San Diego, CA: Academic.
- Sue, D. (2010). *Microragresssions in everyday life: race gender, and sexual orientation*. Hoboken, N.J.: Wiley and Sons.
- Tajfel, H., & Turner, J.C. (1986). The social identity theory of intergroup behavior. In S. Worchel & G. Austin (Eds.), *Psychology of intergroup relations* (pgs. 7–24). Chicago: Nelson-Hall.
- Twenge, J. M., & Crocker, J. (2002). Race and self-esteem: Meta-analyses comparing Whites, Blacks, Hispanics, Asians, and American Indians and comment on Gray, Little, and Hafdahl (2000). *Psychological Bulletin*, 128, 371–408.
- Walton, G. M., & Cohen, G. L. (2007). A question of belonging: race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92, 82–96.

Walton, G., & Cohen, G. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science*, 331, 1447–1451. doi:10.1126/science.1198364

Walton G.M. & Spencer S.J. (2009). Latent Ability: Grades and Test Scores systematically underestimate the Intellectual Ability of Negatively Stereotyped Students. *Psychological Science*, 20, 1132-1139.

Whitt, E.J., Edison, M.I., Pascarella, E.T., Terenzini, P.T., & Nora, A. (2001). Influences on students' openness to diversity and challenge in the second and third years of college. *Journal of Higher Education*, 72 (2), 172-204.

Table 1

[\[back to text\]](#)

List of 26 Effectiveness Factors

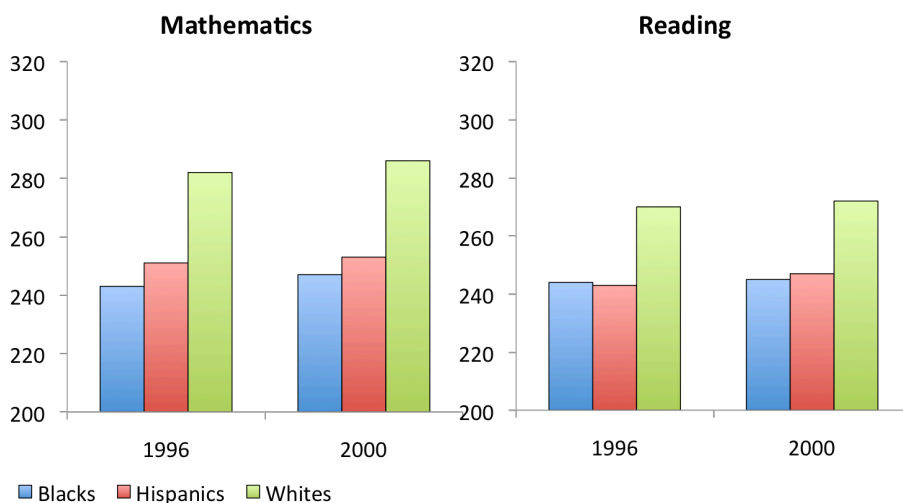
Factors Identified as Important to Lawyer Effectiveness

Analysis and Reasoning
Creativity/Innovation
Problem Solving
Practical Judgment
Researching the Law
Fact Finding
Questioning and Interviewing
Influencing and Advocating
Writing
Speaking
Listening
Strategic Planning
Organizing and Managing One's Own Work
Organizing and Managing Others (Staff/Colleagues)
Negotiation Skills
Able to See the World Through the Eyes of Others
Networking and Business Development
Providing Advice & Counsel & Building Relationships with Clients
Developing Relationships within the Legal Profession
Evaluation, Development, and Mentoring
Passion and Engagement
Diligence
Integrity/Honesty
Stress Management
Community Involvement and Service
Self-Development

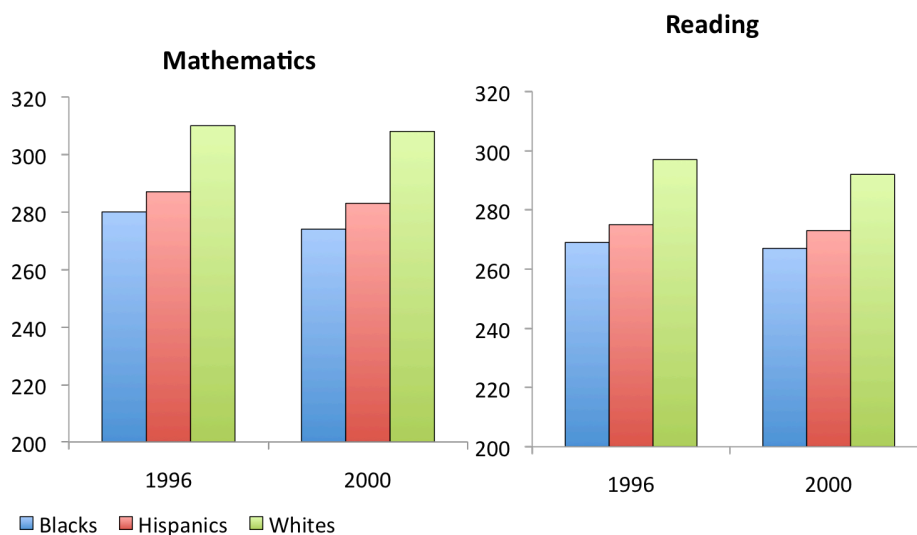
Note: Adapted from Shultz and Zedeck (2003).

Figure 1[\[back to text\]](#)

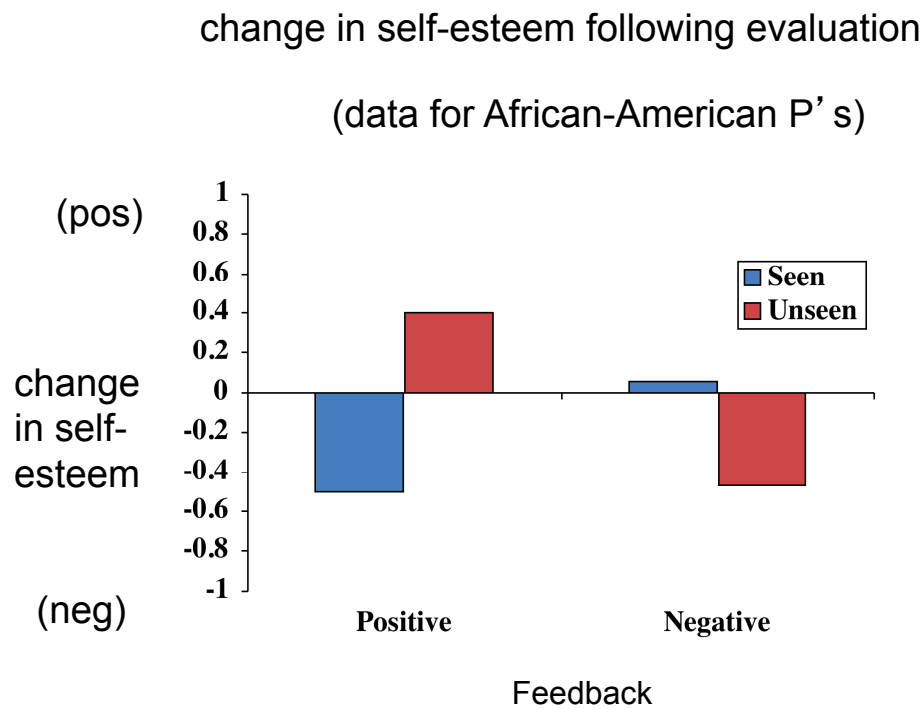
National Assessment of Educational Progress (NAEP) scores, 8th grade



National Assessment of Educational Progress (NAEP) scores, 12th grade



*Bennett, Bridgall, Cauce, Everson, Gordon, Lee,
Mendoza-Denton, Renzulli, & Stewart (2004)*

Figure 2[\[back to text\]](#)

Crocker, Voelkl, Testa, & Major, 1991

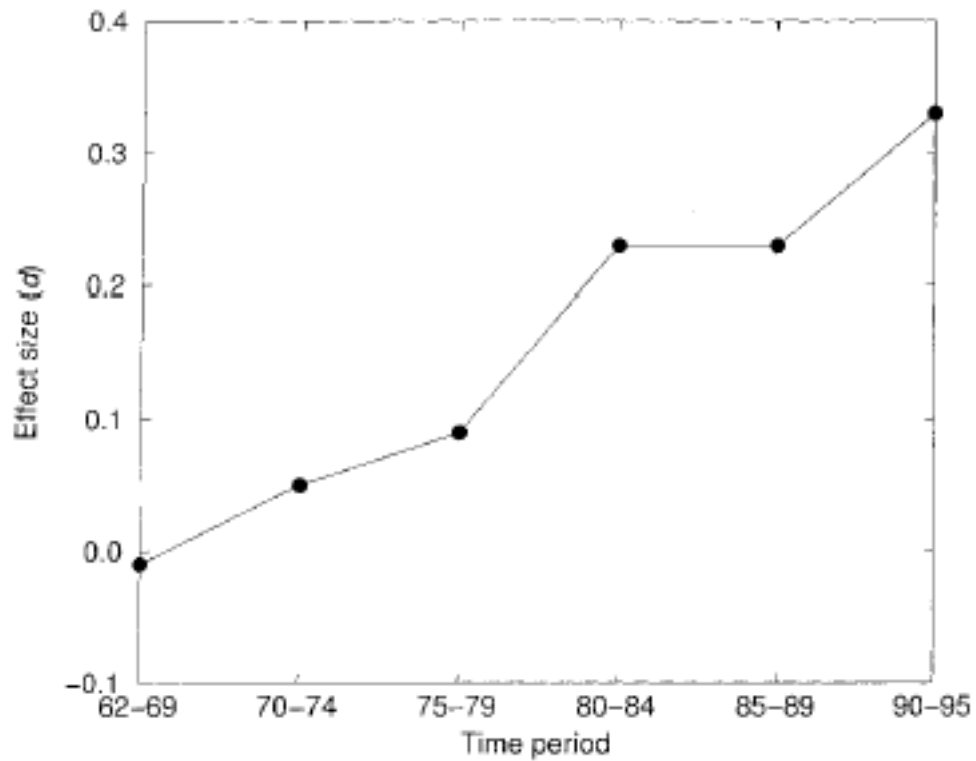
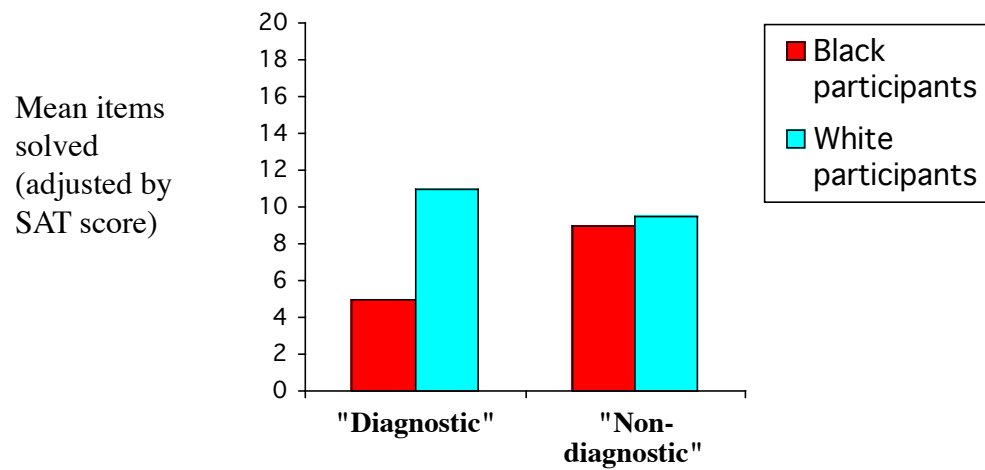
Figure 3[\[back to text\]](#)

Figure 2. Black-White effect size over time. Time periods are given by 20th century calendar year.

Source: Twenge & Crocker (2002)

Figure 4[\[back to text\]](#)

Mean performance on a difficult verbal test as a function of race and test characterization



Steele & Aronson (1995)

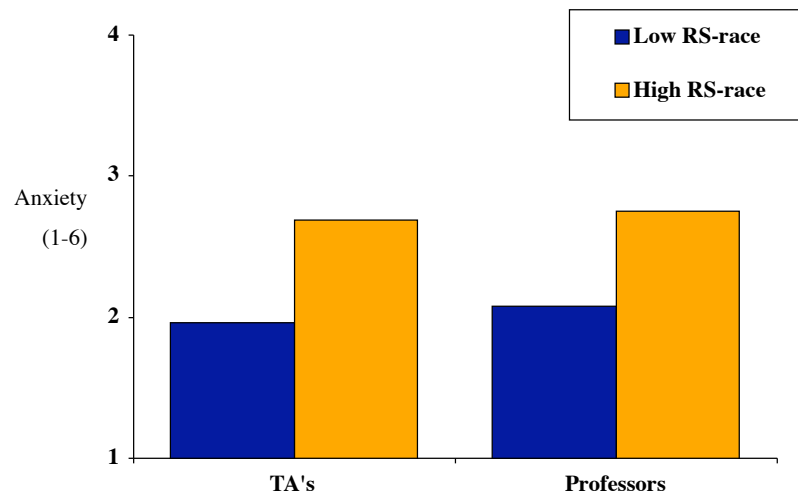
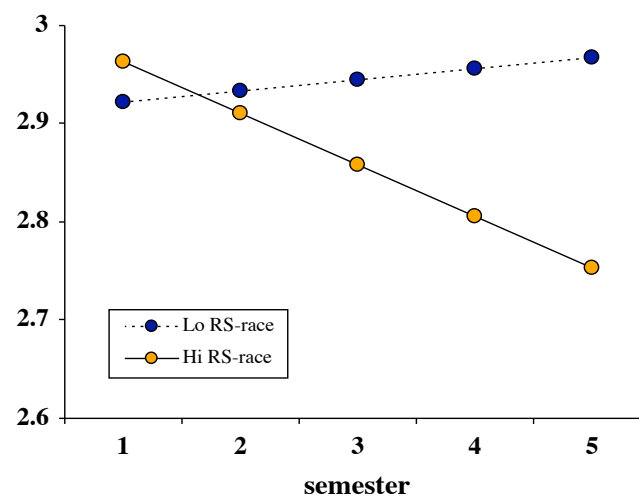
Figure 5

[\[back to text\]](#)

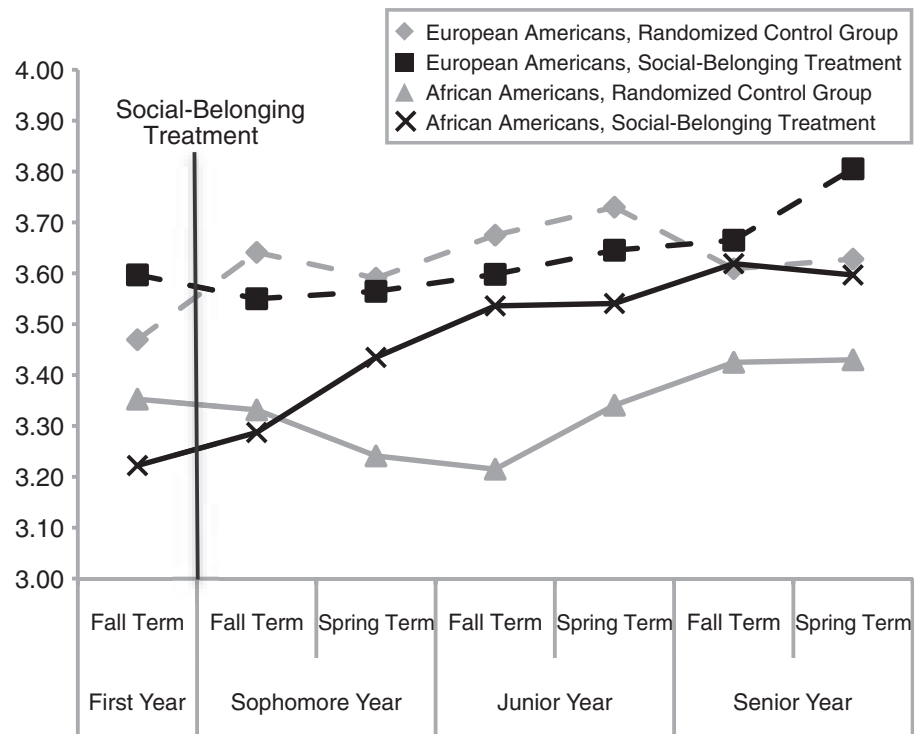
Student protests at the University of California, Berkeley in response to the 2011 “Diversity Bake Sale”



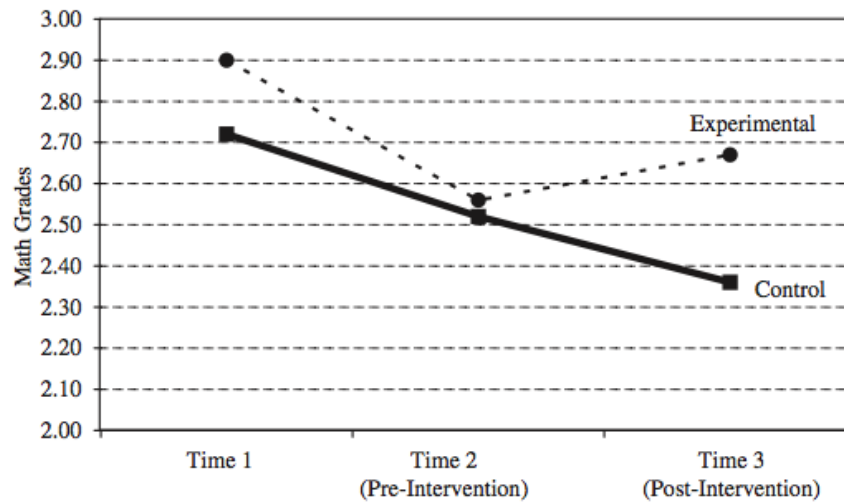
(Photo: Author)

Figure 6[\[back to text\]](#)**Anxiety about asking for help with an academic problem****Change in GPA**

Mendoza-Denton, Downey, Purdie, Davis, & Pietrzak, 2002

Figure 7[\[back to text\]](#)

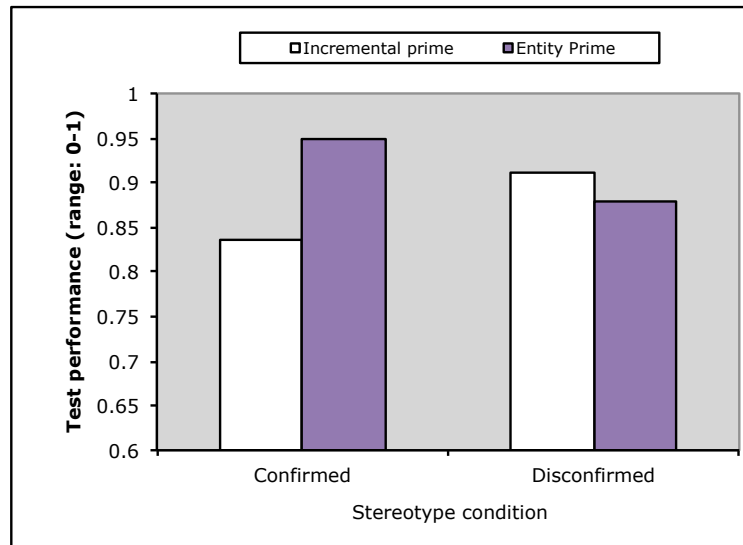
Walton & Cohen, 2011

Figure 8[\[back to text\]](#)

Blackwell, Trzesniewski, and Dweck (2007)

Figure 9[\[back to text\]](#)

Asian background participants:
Mean performance on math test as a function of stereotype
condition and theory prime



Mendoza-Denton, Kahn, & Chan, 2008