

Michigan Journal of Race and Law

Volume 5

1999

Expert Report of Claude M. Steele

Claude M. Steele
Stanford University

Follow this and additional works at: <https://repository.law.umich.edu/mjrl>



Part of the [Civil Rights and Discrimination Commons](#), [Education Law Commons](#), [Law and Race Commons](#), [Law and Society Commons](#), and the [Legal Education Commons](#)

Recommended Citation

Claude M. Steele, *Expert Report of Claude M. Steele*, 5 MICH. J. RACE & L. 439 (1999).
Available at: <https://repository.law.umich.edu/mjrl/vol5/iss1/18>

This Feature is brought to you for free and open access by the Journals at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Michigan Journal of Race and Law by an authorized editor of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.

EXPERT REPORT OF CLAUDE M. STEELE

Gratz, et al. v. Bollinger, et al., No. 97-75321 (E.D. Mich.)
Grutter, et al. v. Bollinger, et al., No. 97-75928 (E.D. Mich.)

I have been Chair of the Department of Psychology at Stanford University since 1997, and a Professor of Psychology since 1991. Prior to that, I was a Professor of Psychology at the University of Michigan from 1987 to 1991; during the last two years at the University of Michigan, I also served as a Research Scientist for the Institute for Social Research. Before that, I was a member of the faculty at the University of Washington from 1973 to 1987. I have written extensively about the psychology of how minority groups, especially African Americans, contend with negative stereotypes and the role this process can play in their school achievement and standardized test performance. A complete curriculum vitae, including a list of publications, is attached hereto as Appendix A.[†] I have not testified as an expert at trial or by deposition in any prior case. I am being compensated at a rate of \$200 per hour for my work in connection with this matter.

My testimony is based, most generally, on an expertise that has been developed over a 25-year period of research in the areas of social psychology, the social psychology of race and race relations, and the effects of race on standardized test performance. In preparing this testimony I have consulted a broad range of knowledgeable colleagues and experts in these areas, as well as the relevant research literature. My testimony is also based on a 10-year research program that I have directed, the aim of which has been to understand the role of race and gender stereotypes in shaping test performance and the formation of academic identities.

Although most of the relevant data used in this report comes from research done on the SAT exam (the Educational Testing Service has broadly disseminated substantial information on the characteristics and validity of data on the SAT), my conclusions can fairly be generalized to the ACT and LSAT exams, as well. These tests are so similar in the way they are constructed, what they measure, and their purpose (aids to admission decisions in higher education) that I treat them as a single class of tests with the presumption that, as far as my testimony goes, what is said of one test generalizes to the others as well. Throughout my testimony, then, when reference is made to testing data, unless otherwise specified, it refers to data based on the SAT.

† Appendix A has not been reproduced here.

OPINIONS TO BE EXPRESSED

Standardized admissions tests such as the SAT, the ACT, and the LSAT are of limited value in evaluating “merit” or determining admissions qualifications of all students, but particularly for African American, Hispanic, and American Indian applicants for whom systematic influences make these tests even less diagnostic of their scholastic potential. The first part of this caution—that the test should not be relied upon too heavily in general admissions—is a standard recommendation of the companies that produce these tests, but is also based on extensive evidence documenting the limited predictiveness of these tests. This is not surprising given that these tests are not designed to measure innate ability nor mastery of a specified curriculum. Instead, standardized tests measure developed skills.

The second part of the caution with respect to standardized tests—that use of these tests with minority applicants is especially unreliable—is based on longstanding research, including work done in my own laboratory over the past 10 years, showing that experiences tied to one’s racial and ethnic identity can artificially depress standardized test performance. Importantly, these effects go beyond any effects of socio-economic disadvantage, affecting even the best prepared, most invested students from these groups who often come from middle-class backgrounds. Relying on these tests too extensively in the admissions process will preempt the admission of a significant portion of highly qualified minority students. In making this argument, I will address three issues: The nature of the mental capacity measured by these tests; how well these tests predict performance in higher education for all students; and reasons African American, Hispanic, and American Indian students are more likely to underperform on these tests.

I. WHAT KIND OF CAPACITY IS MEASURED BY
STANDARDIZED ADMISSIONS TESTS?

1. *How are the SAT, ACT, and LSAT designed?* To understand what these tests do and do not measure, it is important first to understand how they are constructed. In the first step, a group of professional item writers and content area experts generate a large pool of test items in the areas covered by the test. In this process, the test makers are guided by general guidelines about what skills and knowledge are critical to succeeding in a given area. But these guidelines are not derived from some clearly specified theory or knowledge of how to measure intelligence or scholastic aptitude in these areas. They are settled on, for the most part, by con-

sensus among the item generators and the board of area experts who they consult.

Next, these items are given to a norming sample of people who are selected for either being a representative or a random sample of the population for whom the test is to be used. Roughly speaking, items that correlate with school grades in this norming sample are kept on the test and items that do not correlate well with grades in this sample are dropped from the test. For example, correct answers given on test items involving algebra by a student who received high grades in his or her algebra classes, would be kept because they correlated positively with school success. In this way, items are identified that, for this population, are associated with school success, or in testing parlance, are “predictive” of school success. The resulting test can then be administered in this population with the feature that one’s score on it will be somewhat predictive of the grades one will achieve. Like most standardized scholastic tests, the SAT, ACT, and LSAT are all constructed in this way.

2. *What do these tests measure?* The overriding implication of this construction procedure is that it is difficult to answer this question with a precise, conceptual definition. As has been classically said, “scholastic aptitude is what scholastic aptitude tests measure.” The content of the test is not derived from a clear conception of the aptitude under test, and the inclusion of items on the test is decided empirically—by which items correlate with school grades in the norming sample. To develop a conceptual understanding of the mental capacities measured by the test, one would have to do what test researchers do: Work backwards by trying to discern through factor analysis of the items selected what underlying capacities they measure.

Two things about the nature of these tests that bear on their use in college and law school admissions can be said with certainty. First, based on this test construction methodology it is clear that the items on these tests measure what has to be substantially learned or “developed” skills and knowledge. Many factors including heredity may underlie scholastic aptitude, but even the highest estimates of hereditary influence allow for substantial influence of experiential factors. This means that one’s performance on these tests can be influenced by one’s experience, by one’s cultural background, by one’s access to schooling and the cultural perspectives, attitudes, and know-hows that might favor test performance, by the extent to which one’s peers value school achievement, by the nature of one’s dinner table conversation, and so on. This point will be important to my later discussion of the role of race and ethnicity in influencing performance on these tests. In addressing those issues, it is important to emphasize that the SAT, ACT, and LSAT are not tests of

innate ability that are impervious to experiential influences. Quite the opposite is true.

The second point about test content that can be made with certainty is that, in addition to not measuring mental capacity, neither are they achievement tests: they are not constructed to test how much one has learned from a specifiable curriculum. Rather, they are described by their makers as “aptitude” tests. I have just explained how difficult it is to conceptually define the “aptitude” they measure (other than to say that it is a measure of test-taking aptitude). But it is not the case that, not measuring a specifiable aptitude, they do measure achievement or how much one has learned in school. Ours is the only nation in the world that uses aptitude tests in higher education admissions rather than tests that measure achievement—how much a person has learned in earlier schooling, which are typically better predictors of success in higher education than aptitude tests.

In sum, then, as the companies that make them acknowledge, the SAT, ACT, and LSAT measure a set of scholastic skills that are neither innate nor directly influenced by school curricula. Thus the value of these tests in informing admissions decisions depends not on assessing some well-defined talent or knowledge base, but solely on their empirically determined ability to predict college or law school grades. How well, then, do they predict these grades?

II. HOW GOOD ARE STANDARDIZED ADMISSION TESTS AT PREDICTING SUCCESS IN HIGHER EDUCATION?

The SAT is popularly assumed to measure such a singularly important component of academic merit as to mandate its centrality in the admissions process. Among the most common rationales for using it to make admissions decisions, in addition to the use of school grades, is that it taps a form of scholastic aptitude that is not dependent on the quality of one’s high school curriculum—thus the idea that it measures an underlying, if not innate, aptitude. In contrast to most people’s expectations, however, the SAT in fact measures only about 18% of the factors that determine a person’s freshman grades. And this figure holds even when controlling for the difficulty of the courses taken. (It also holds when the statistical problem of restriction of range is controlled for.) Moreover, the SAT adds hardly any predictive power in the prediction of freshman grades over what one gets from using high school grades alone. That is, using the SAT only increases one’s prediction of freshman grades by about 3% or 4% over what one could predict using high school grades alone. And as the criterion measures get farther away in time from when

the SAT is taken—as for sophomore grades, graduation rates, and professional success—the correlations with the SAT get substantially smaller.

An important implication of this fact is that even large score differences on the SAT do not translate into very large differences in the skills that underlie grade performance. This is what is implied by the small relationship between scores on the test and subsequent grades: that relatively few of the skills critical to grades are measured by the tests. And this, in turn, means that a score difference between two people, or between two groups (for example, Blacks and Whites), that is as large as say, 300 points, a difference that can sound big, actually represents a very small difference in skills critical to grade performance.

Perhaps the limitations on the usefulness of these tests can be made clearer with an analogy. Suppose that you were confined to selecting a basketball team based on how many of 10 free throws a player hits. The first thing you'd worry about is having to select basketball players based on the single criterion of free throw shooting, which you know is only a small portion of the skills that go into actual basketball playing. Even worse, you would know that you would never pick Shaquille O'Neal. Similarly, standardized tests tap only a small set of the skills that make a good student—approximately the 18% that I mentioned.

Another problem you would have selecting your basketball team would be how to interpret a player's scores. If a player hits 10 of 10 or 0 of 10 you would be fairly confident about making a judgment; the 10 of 10 guy you keep, the 0 of 10 guy you drop. But what about the player who hits 3, 4, 5, 6, or even 7? Middling scores like these could be influenced by many things other than underlying potential for free throw shooting or basketball playing, such as the amount of practice involved, access to effective coaching, whether the player was having a good or a bad day. Roughly the same is true, I suggest, for interpreting standardized test scores: Extreme scores (though less reliable) might permit some confidence in a student's likelihood of success, but middling scores are more difficult to interpret as an indication of underlying promise. Are they inflated by middle-class advantages such as prep classes, private schools, and European Cathedral tours? Or are they deflated by race-linked experiences such as social segregation and being consistently assigned to the lower tracks in school?

Although test scores can be useful and do have the ability, however limited, to inform admission decisions, the fact is that they simply do not capture any large portion of what makes up academic potential or merit. Grades depend on many things not measured by these tests, and admissions committees should use them with caution and only together with as much other information about candidates as can be obtained. This advice

holds for students from any background. But there are reasons to believe that this advice is especially important in the case of minorities.

III. ARE THERE SIGNIFICANT FACTORS THAT MIGHT CAUSE AFRICAN AMERICAN, HISPANIC, AND AMERICAN INDIAN STUDENTS TO PERFORM LESS WELL THAN OTHER GROUPS ON THESE TESTS?

The answer to this question is a resounding, "Yes." I describe here what I regard as the two most important such factors.

Stereotype threat and test performance. My research, and that of my colleagues, has isolated a factor that can depress the standardized test performance of minority students—a factor we call stereotype threat. This refers to the experience of being in a situation where one recognizes that a negative stereotype about one's group is applicable to oneself. When this happens, one knows that one could be judged or treated in terms of that stereotype, or that one could inadvertently do something that would confirm it. In situations where one cares very much about one's performance or related outcomes—as in the case of serious students taking the SAT—this threat of being negatively stereotyped can be upsetting and distracting. Our research confirms that when this threat occurs in the midst of taking a high stakes standardized test, it directly interferes with performance.

In matters of race we often assume that once a situation is objectively the same for different groups, that it is *experienced* the same by each group. This assumption might seem especially reasonable in the case of "standardized" cognitive tests. But for Black students, unlike White students, the experience of difficulty on the test makes the negative stereotype about their group relevant as an interpretation of their performance, and of them. Thus they know as they meet frustration that they are especially likely to be seen through the lens of the stereotype as having limited ability. For those Black students who care very much about performing well, this is an extra intimidation not experienced by groups not stereotyped in this way. And it is a serious intimidation, implying, as it does, that they may not belong in walks of life where the tested abilities are important walks of life in which they are heavily invested. Like many pressures, it may not be fully conscious, but it may be enough to impair their best thinking.

To test this idea, Joshua Aronson and I asked Black and White Stanford students into our laboratory and, one at a time, gave them a very difficult 30-minute verbal test, the items of which came from the advanced Graduate Record Examination in literature. The bulk of these students were sophomores, which meant that the test would be difficult for them—precisely the feature that we reasoned would make this simple

testing situation different for our Black participants than for our White participants. We told each student that we were testing ability.

Black students performed dramatically worse than White students on the test. As we had statistically equated both groups on ability level, the differences in performance were not because the Black students had weaker skills than the White students. Something else was involved. Before we could confirm that that "something else" was stereotype threat, we had to control for the possibility that the Black students performed worse than the White students because they were less motivated or because their skills could be somehow less easily extrapolated to the advanced material of this test. We concluded that if stereotype threat and not something about these students themselves had caused their poorer test performance, then doing something that would reduce this threat during the test should allow their performance to improve, to go up to the level of equally capable White students. We devised a simple way to test this: We presented another group of Black and White sophomores, again statistically equated on ability level, the same test we had used before—not as a test of ability, but as a "problem-solving" task that had nothing to do with ability. This made the stereotype about Blacks' ability irrelevant to their performance on the task since, ostensibly, the task did not measure ability. A simple instruction, yes, but it profoundly changed the meaning of the situation. It told Black participants that the racial stereotype about their ability was irrelevant to their performance on this particular task. In the stroke of an instruction, the "stereotype spotlight," as psychologist Bill Cross once called it, was turned off.

As a result, Black students' performance on this test matched the performance of equally qualified Whites. With the stereotype spotlight on, Blacks performed dramatically worse than Whites; with it off, they performed the same. Thus, stereotype threat of the sort that we argue characterizes the daily experiences of Black students on predominantly White campuses and in a predominantly White society, can directly affect important intellectual performances such as standardized test performance.

But it has broader effects too. Stereotype threat follows its targets onto campus, affecting behaviors of theirs that are as varied as participating in class, seeking help from faculty, contact with students in other groups, and so on. And as it becomes a chronic feature of one's school environment, it can cause what we have called "disidentification"; the realignment of one's self-concept and values so that one's self-regard no longer depends on how well one does in that environment. Disidentification relieves the pain of stereotype threat by breaking identification with the part of life where the pain occurs, which necessarily includes a loss of motivation to succeed in that part of life. When school is the part of life where stereotype threat is felt—as for women in advanced math or

African Americans in all areas—disidentification can be a costly and life-altering adaptation.

In subsequent years, our research has revealed several important parameters of the effect of stereotype threat on standardized test performance. First, it can interfere with the test performance of any group whose abilities are negatively stereotyped in the larger society: Women taking difficult math tests; lower-class French students taking a difficult language exam; older people taking a difficult memory test; White male athletes being given a test of natural athletic ability; White males taking a difficult math test on which they are told “Asians do better”; as well as Hispanic students at the University of Texas being given a difficult English test. This research shows stereotype threat to be a very general effect, one that is undoubtedly capable of undermining the standardized test performance of any group negatively stereotyped in the area of achievement tested by the test.

We have also discovered that the detrimental effect of stereotype threat on test performance is greatest for those students who are the most invested in doing well on the test. As an intimidation, one might expect that it would affect the weakest students most. But this is not what happens. Across our research, stereotype threat most impaired students who were the most identified with achievement, those who were also the most skilled, motivated, and confident—the academic vanguard of the group more than the academic rearguard.

This fact had been beneath our noses all along in our data and even in our theory. A person has to care about a domain in order to be disturbed by the prospect of being stereotyped in it. So all of our earlier experiments had selected participants who were identified with the domain of the test involved—Black students identified with verbal skills and women identified with math. But we had not tested participants who were less identified with these domains. When we did, what had been beneath our noses hit us in the face. None of these disidentified students showed any effect of stereotype threat whatsoever. Nothing.

Now make no mistake, these disidentified students did not perform well on the tests. Like anyone who does not care, they would start the test, discover its difficulty, stop trying very hard and get a lower score. But their performance did not differ depending on whether they were at risk of being judged stereotypically—their performance was the same regardless of whether they had been told it was their ability we were testing.

This finding tells us two important things. The first is that the poorer standardized test performance of Black students may have two sources. One is more commonly understood: It is the poorer performance of some among this group who are not well prepared and perhaps

not well identified with school achievement. The other, however, has not been well understood: The underperformance among strong, school-identified members of this group whose lower performance reflects the stereotype threat they are under.

But these findings make a point of some poignance as well: The characteristics that expose this vanguard to the pressure of stereotype threat is not weaker academic identity and skills, but stronger academic identity and skills. They have long seen themselves as good students, better than most other people. But led into the domain by their strengths, they pay an extra tax on their investment there, a "pioneer tax," if you will, of worry and vigilance that their futures will be compromised by the ways society perceives and treats their group. And it is paid everyday, in every stereotype-relevant situation. Recent research from our laboratory shows that this tax has a physiological cost. Black students performing a cognitive task under stereotype threat had elevated blood pressure.

This finding raises another point: Being a minority student from the middle-class is no escape from stereotype threat and its effect on standardized test performance or performance in higher education more generally. In the American mind we have come to view the disadvantages associated with being Black, for example, as disadvantages of social and economic resources and opportunity. This assumption is often taken to imply its obverse: That is, if you are Black and come from a home that has achieved middle-class status, your experiences and perspectives are no longer significantly affected by race. Our research shows quite clearly that this is not so. In fact, if being middle-class gave you the resources that helped you identify with school achievement, ironically, it may lead you to experience stereotype threat even more keenly. It is investment in the domain of schooling—often aided by the best resources and wishes of middle-class parents—that can make one, at the point of reaching the difficult items on the SAT, experience the distracting alarm of stereotype threat.

All of these findings then, taken together, constitute a powerful reason for treating standardized tests as having limited utility as a measure of academic potential of students from these groups. But there are other reasons as well.

Different experiences. The point here is that factors like race, social class, and ethnicity still shape the life trajectories and experiences of individuals in society and as a result, can have profound effects on test performance. For example, consider what being African American, even from the middle-class, can predispose a person to experience: Assignment to lower academic tracks throughout schooling; being taught and counseled with lower expectations by less skilled teachers in more poorly funded schools; attending school in more distressed neighborhoods or in

suburban areas where they are often a small, socially isolated minority; living in families with fewer resources; and having peers who—alienated by these conditions—may be more often disinterested in school. Clearly these race-linked experiences are enough to lead students from this group to have lower scores on the SAT at the point of applying to college without any reference to innate ability. A similar scenario could be described for many Hispanic groups in this society and for American Indians (especially those living on reservations).

If one thinks of all the relationships, experiences, and motivations that underlie good test performance as a river or confluence of influences, it is clear that some groups will have more access to this river than others. Accordingly, those with less access, by dint of the weaker academic and test performance skills this causes, will have lower test scores and thus more limited access to higher education. Of course, to the extent that the skills they lack are critical to success in school, this limitation of access is appropriate under the ideal of sending the most qualified students on to higher education. But it is important to stress, even here, that for these students, their lower test scores may reflect their limited access to the critical confluence of experiences as much as any real limitation in potential for higher education.

Again the free-throw analogy might be helpful. The part of this analogy most relevant to the present point is how to interpret the performance of people who, for sociocultural reasons, have had little exposure to free-throw shooting. They are not likely to hit many shots. But the problem is how to interpret their poor performance vis a vis their potential to play basketball. Their poor free-throw shooting could reflect problems that would make them very poor basketball players, or it could reflect a lack of experience that could be easily overcome, or even an orientation that while hurting free-throw shooting might *help* basketball playing. It would be difficult to know. And this is the fundamental ambiguity surrounding the interpretation of low SAT scores among students from backgrounds without significant access to the culture represented on the test. Their lower scores are more difficult to interpret.

CONCLUSION

In recent years the media has made a great deal of the fact that minority students on a college campus often have lower average SAT scores than Whites and Asians on the same campus. The clear implication, presumably taken up by the public, is that SAT gaps of this size reflect that the minorities being admitted are “less qualified” than the White and Asian students. My testimony, I hope, has put these gaps in a different light: Gaps of this size actually represent only a tiny difference in the real

skills needed to get good college or law school grades and they reflect the influence of a complex of factors tied to race in our society that, for reasons unrelated to real academic potential, depress minority student test scores. Furthermore, this gap is almost never caused by there being a lower admissions threshold for Blacks than for Whites or Asians. It reflects the fact that there is a smaller proportion of Black than Whites and Asians with very high SAT scores. Thus, when you average each group's scores, the Black average will be lower than the White and Asian averages. Why there is a smaller proportion of Blacks with very high scores is, of course, a complex question with multiple answers involving, among other things, the effects of race on educational access and experience, as well as the processes dwelt on in this document. The point, though, is that Black test score deficits are taken as a sign of their being underprepared when, in fact, virtually all Black students on a given campus have tested skills completely "above threshold" within the range of the tested skills for other students on the campus, and in this sense, have skills up to the competition.

Having made these arguments, I hope to have provided a better understanding of minority students' underperformance on standardized tests and of what that underperformance means with regard to their ability to succeed in higher education. It is simply the case that we have no single, or even small, set of indicators that satisfactorily captures "merit" or "potential" for academic success and a contributing life.

REFERENCES CONSULTED:

Jencks, C., & Phillips, M. (1998). *The Black-White Test Score Gap*. Washington, DC.: Brookings Institution Press.

Lemann, N. (1997, September). The great sorting. *The Atlantic*.

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1994). Student group differences in predicting college grades: sex, language, and ethnic groups. (College Board Report No. 93-1, ETS No. 94.27). New York: College Entrance Examination Board.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist* 52. 613-619.

Steele, C. M. (1998). Thin ice: On being African American in college. Manuscript under review.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*. 69. 797-811.

Sturm, S., & Guinier, L. (1996). The Future of Affirmative Action: Reclaiming the Innovative Ideal. 84 Cal. L. Rev. 953-1036.

Wightman, L. (1997). The Threat to Diversity in Legal Education: An Empirical Analysis of the Consequences of Abandoning Race as a Factor in Law School Admissions Decisions. 72 N.Y.U. L. Rev. 1.