

84 Calif. L. Rev. 953, *

Copyright (c) 1996 California Law Review
California Law Review

July, 1996

84 Calif. L. Rev. 953

SYMPOSIUM: RACE-BASED REMEDIES: Rethinking the Process of Classification and Evaluation: The Future Of Affirmative Action: Reclaiming The Innovative Ideal

Susan Sturm *, Lani Guinier **

SUMMARY:

... We are witnessing a broad-based assault on affirmative action--in the courts, the legislatures, and the media. Opponents have defined affirmative action as a program of racial preferences that threatens fundamental American values of fairness, equality, and democratic opportunity. ... Rethinking our assumptions about selection is important to be able to pursue goals of racial and gender justice and fairness. ... Fairness means that applicants should be numerically ranked for selection using "objective," race- and gender-neutral selection criteria. ... This heightened visibility of race- and gender-driven exceptions to "objectively derived" test-score-based selection plays into existing biases and stereotypes, particularly about race. ... Thus, visibility, combined with underlying racism or sexism, undermines the capacity of non-beneficiaries of affirmative action to see race- or gender-linked departures from selection criteria as fair. ... For example, in response to evidence that the old-boy-network approach to recruitment excluded people of color and women, many firms and schools have abandoned informal networking in favor of advertising, posting, and active searches, which expand the applicant pool. ...

[*953]

We are witnessing a broad-based assault on affirmative action--in the courts,¹ the legislatures,² and the media.³ Opponents have defined affirmative action as a program of racial preferences that threatens fundamental American values of fairness, equality, and democratic opportunity.⁴ Opponents successfully depict racial preferences as **[*954]** extraordinary, special, and deviant--a departure from prevailing modes of selection. They also proceed on the assumption that, except for racial or gender preferences, the process of selection for employment or educational opportunity is fair, meritocratic, and functional.⁵ Thus, they have positioned affirmative action as unnecessary, unfair, and even un-American.⁶

Those of us pursuing the quest of racial and gender justice in a genuinely democratic society face a crucial challenge. How do we respond to this assault

on affirmative action? How do we invite a deeper conversation and analysis of selection and admissions conventions in pursuit of fairness? Understandably, much of the response has been reactive. Supporters of affirmative action typically engage the debate on the terms defined by the assault: affirmative action must continue. It is fair. It is still needed to rectify continued exclusion and marginalization in the society. ⁷

Supporters of affirmative action have also put forward a critique of the fairness and functionality of existing merit standards. ⁸ They mar- **[*955]** shal considerable evidence showing that these standards exclude women and people of color, and that people who were excluded in the past do not yet operate on a level playing field. ⁹ They have also challenged the justification for relying on these exclusionary criteria; they argue that the selection criteria do not predict the future performance of candidates in the positions they seek to occupy. ¹⁰ They then rely on this critique of the fairness and validity of existing merit standards to justify departing from those standards for women and people of color. Affirmative action is justified to level the playing field, to rectify the biases built into the existing selection system, and to remedy past and continuing exclusion or underrepresentation. ¹¹

Despite the moral and empirical force of these arguments, there is a sense in which they are not being heard. They certainly have not reshaped the terms or tone of the public debate. The most compelling moral claims are simply dismissed as special-interest pleading. Part of the reason for this asymmetry is that proponents of racial and gender justice have responded to the debate only as it is framed by the current assault.

This narrow response has tactical, strategic, and substantive costs. As a tactical matter, proponents have accepted a paradigm that misdirects attention and energy into trench warfare, rather than into pursuing a progressive agenda. By reacting defensively to the current onslaught, they have foreclosed discussion of new, innovative strategies for racial and gender justice. Substantively, they have accepted an existing **[*956]** framework of selection that is fundamentally and deeply flawed for those whom it includes as well as for those left out.

In other words, affirmative action, as it is currently practiced, supplements an underlying framework of selection that is implicitly arbitrary and exclusionary. It does not challenge the overall operation of a conventional and static selection process; instead, it creates exceptions to that process. Those exceptions play into existing racial stereotypes, predictably generating backlash. ¹² By implicitly legitimizing a selection process that operates in the name of merit, affirmative action programs reinforce that backlash. Programs perceived as racial preferences also enable employers to cast issues of economic retrenchment in terms of racial conflict. Many white workers who acknowledge the lack of corporate responsibility for the economic well-being of workers still focus their wrath and blame on the workers perceived as beneficiaries of affirmative action. ¹³

It is time, we argue, for those of us committed to racial and gender equity to advance a more fundamental critique of existing selection and admission conventions. It is time to discuss how conventional assessment and predictive criteria do not function fairly, democratically, or even meritocratically for many

Americans who are not members of racial or gender minorities. To reclaim the moral high ground, we must broaden and expand the terms of engagement. By revealing faulty assumptions about the concept of affirmative action and the system of selection in which it operates, we can move from an incrementalist strategy of inclusion for a few to a transformative vision of reform for the many.

To reopen the conversation on race, gender, and democratic opportunity, it is necessary to change the paradigm. Certainly, we must challenge out loud the basic assumption that affirmative action is a departure from an otherwise sound meritocracy. At the same time, we must challenge existing add-on practices of affirmative action as too conservative a remedy. The experience of women and people of color [***957**] offers insights beyond showing how and why those particular people have been excluded. We need to show that the current one-size-fits-all ranking system of predicting "merit" is no longer justified or productive for anyone.

The present system of selection is unfair for people who are neither women nor people of color. It denies opportunity for advancement to many poor and working-class Americans of all colors and genders who could otherwise obtain educational competence. It is underinclusive of those who can actually do the job. It is deeply problematic as a predictor of actual job performance. Across-the-board, it does violence to fundamental principles of equity and "functional merit" ¹⁴ in its distribution of opportunities for admission to higher education, entry-level hiring, and job promotion.

Typical among the existing criteria and selection methods are paper-and-pencil tests, such as the Scholastic Assessment Test (SAT), the Law School Admissions Test (LSAT), and civil service exams. ¹⁵ These tests, which are used to predict future performance based on existing capacity or ability, do not correlate with future performance for most applicants, at least not as a method of ranking those "most qualified." ¹⁶ These tests and informal criteria making up our "meritocracy" tell us more about past opportunity than about future accomplishments on the job or in the classroom. ¹⁷

In challenging the way these tests are used, we are not proposing a critique of merit per se. Nor are we advancing an entirely original argument. Simply stated, we seek to highlight the way that certain paper-and-pencil tests have been used as "wealth preferences" or poll taxes to determine who gets to participate as full citizens in our democracy. ¹⁸ As Michael Lind argues in a slightly different context, these tests are used, in conjunction with subjective assessments and informal networks, to develop a class-linked opportunity structure that credentializes "a social oligarchy." ¹⁹

The approach we develop in this Article links affirmative action initiatives with the project of fundamentally rethinking how we define and practice genuine merit selection. We argue that affirmative action is [***958**] an opportunity to take from the margin to rethink the whole. Affirmative action is not about exceptions to the norm; it is about the norm itself. Affirmative action, and the experience of those who have been previously excluded, provide a window on a much larger set of questions. These are the same questions that companies and educational institutions must face to meet the demands of an economy in transition: Can we define and predict ability to perform based on one-size-fits-

all tests and criteria? How do we go about identifying the type of worker/student who will perform successfully under changing economic conditions? Is sameness fairness? Or must we reconsider the notion that in a complicated world there are simple and single solutions? How do we rethink the process and content of selection to better accommodate the demands of the twenty-first-century workplace?

It is time to ask a different set of questions about affirmative action, questions that address the most pressing problems facing not only people of color and women, but all of those who are unfairly excluded from participation in work and education. We need to go beyond the modest curative of affirmative action to examine more deeply our system of selecting and evaluating all workers and students. This approach to affirmative action can open up an inquiry into the adequacy and legitimacy of the one-size-fits-all approach to selection that prevails in many arenas. In this way, affirmative action provides a less reactive, more transformative critique that highlights the range of preferences implicit in conventional selection and prediction criteria.

Patterns of exclusion experienced primarily by women and people of color are, nevertheless, still important. They serve as signals. Patterns of race- and gender-based exclusion signal the possibility that bias or unfair advantage has operated in the ostensibly neutral selection process. They also signal the inadequacy of traditional methods of selection for everyone, and the need to rethink the process used to allocate opportunities to participate in work and school. In other words, patterns of exclusion provide a window on the methods for "inclusion." They are an important source of continuous critique of monolithic and monochromatic ranking and selection processes.

Rethinking our assumptions about selection is important to be able to pursue goals of racial and gender justice and fairness. Even more, it is crucial to our capacity to develop productive, fair, and efficient institutions that can meet the challenges of a rapidly changing, unstable, and increasingly complex marketplace. By using the experience of those on the margin to rethink the whole, we may forge a new, progressive vision of cross-racial collaboration, functional diversity, and genuinely democratic opportunity. **[*959]**

Part I of this Article sets out the stock narratives underlying the affirmative action debate and the assumptions about merit and fairness that underlie those narratives. Part II then shows that the current meritocracy is neither fair nor functional, and that it in effect gives preference to candidates who enjoy privileged socio-economic positions. Part III articulates the need for a new paradigm for recruitment, selection, and promotion. The goals of this new framework are threefold: 1) to locate and develop workers who can do the job, 2) to attain genuine inclusion of underrepresented groups, and 3) to promote a collaborative opportunity structure that brings fresh perspectives to doing a better job. Part IV describes one alternative: a framework for selection that shifts the focus from prediction to experience, based on structured, participatory, and accountable assessment.²⁰ In this final Part, we suggest that new paradigms for affirmative action can integrate diversity and merit, and thereby build into the framework of selection the capacity to adapt to the innovative challenges of a dynamic and uncertain economy.

The Stock Affirmative Action Narratives

In this Section we make visible the premises and assumptions that typically frame the affirmative action debate. Many employers and educational institutions select applicants by ranking them through a combination of paper-and-pencil tests and subjective assessments. ²¹ Decision makers equate performance on these "neutral" selection criteria with merit. The assumption is that this system is fair and functional, that applicants have an unbiased opportunity to compete for positions, and that the resultant meritocracy operates to identify the people who are most qualified for and deserving of the position.

We begin by sketching out the prototypical selection processes at issue in the affirmative action debate. We describe how affirmative action typically operates within these scenarios. We also identify the particular conceptions of fairness and merit that permeate the debate, and show that they are both overly narrow and historically recent. This discussion lays the foundation for the subsequent demonstration that these underlying premises about selection are both unfair, in that they arbitrarily exclude some people and advantage others, and invalid, in that they fail to define either the goals or attributes of successful performance or to predict in most cases the individuals who can meet them.

A. The Narratives

Competing narratives drive the affirmative action debate. Each story is propelled by different assumptions about fairness and merit. ²² Each story proceeds from different assumptions about the baseline of decision making: how fair, unbiased, and merit-driven is the system in which affirmative action operates? Although many of these differences in assumptions are never directly expressed, they implicitly provide the analytical framework for the argument advanced by the story. The stock story of affirmative action critics in the employment context (and the one that appears most often in the cases) is of the white civil servant--say a police officer or firefighter--John Doe. He scored several points higher on the civil service exam and interview rating process, but lost out to a woman or person of color who did not score as high on those selection criteria. ²³ John Doe claims, along with many public opponents of affirmative action, that he is more qualified for the job, and that it is unfair to allow race or gender considerations to deprive him of what he "deserves." ²⁴

The recent decision of the United States Court of Appeals for the Fifth Circuit, *Hopwood v. Texas*, exemplifies the stock affirmative action narrative in the education context. ²⁵ The University of Texas Law School based its admissions decisions largely on an applicant's score on the Texas Index ("TI"), a composite of undergraduate grade point average ("GPA") and Law School Aptitude Test ("LSAT") score. The Law School used this number to rank candidates for admission, to predict their performance in the first year of law

school, and to determine the number of offers needed to fill its class.²⁶ To increase the efficiency of the admissions process, the law school sorted the applicant pool into three categories according to applicants' TI scores: presumptive admit, presumptive deny, and a discretionary zone. The law school supplemented this system with an affirmative action program designed to remedy past discrimination in the Texas school system and to increase the diversity of the law school. The law school's affirmative action program used lower ranges to place black and Mexican Americans into the three admissions categories.²⁷

Cheryl Hopwood, a white applicant who was rejected by the law school, scored higher on the TI than some black and Mexican-American applicants who were admitted.²⁸ She, like John Doe, claimed that she was more qualified for admission to the Law School, and that it was unfair to deny her what she "deserved."

Both of these narratives depict the type of affirmative action programs that have come to define the debate: outcome-oriented programs that establish numerical goals to increase the participation of underrepresented groups in various settings.²⁹ These programs establish their goals by determining the percentage of group members (i.e., women, blacks, Latinos, etc.) in the pool from which candidates are drawn. In the educational and employment context, race and gender generally operate as a plus factor in the selection process.³⁰ These programs are frequently referred to, by both supporters and opponents of affirmative action, as racial preferences.³¹ Thus, the stock story frames the affirmative action debate in terms of racial preferences that depart from normal, universal, unbiased, and purportedly fair standards for determining merit. **[*963]**

B. Merit and Fairness in the Stock Affirmative Action Narratives

In essence, John Doe and Cheryl Hopwood are advancing the same two claims: 1) they have more merit than beneficiaries of affirmative action, and 2) as a matter of fairness, they are entitled to the position for which they applied. These two claims rest on the premise that qualifications should determine allocation of employment and educational opportunity, that existing selection criteria determine who is most qualified, and that departures from conventional merit standards are unfair.³²

This argument implicitly defines merit and fairness in particular ways. John Doe and Cheryl Hopwood assume that they "merit" the coveted position.³³ One possible basis for this claim rests on the notion of desert: because they were next in line, based on established criteria of selection, they deserve the position.³⁴ They may also base their claim on the idea of earned recognition: "when an individual has worked hard and succeeded (by her own or others' measures), [she] deserves recognition, praise and/or reward."³⁵ The claim of merit in the stock narrative could also proceed from a functional idea of merit: someone who has the qualities needed to perform effectively in the position under consideration.³⁶ Many affirmative action critics equate functional **[*964]** merit with a numerical ranking on standard paper-and-pencil tests. Those with the higher scores are presumably those who can best

perform in the position under consideration. ³⁷ John Doe and Cheryl Hopwood could also assert that they are better in some abstract or universalistic sense-- that the selection process measures a quality that society generally deems valuable. ³⁸ In conventional terms, those with easily testable or quantifiable types of intelligence are more highly valued and will contribute to the overall quality of the institution, and thus are more qualified. ³⁹

The stock narrative's claim of unfairness builds on these assumptions that merit should and, in the absence of affirmative action, does govern employment and educational decision making. To the extent that affirmative action departs from an otherwise fair and valid system of selection, it is unfair. Fairness, like merit, is also a concept with varying definitions. ⁴⁰ The stock story defines fairness in terms of formal or procedural fairness to the individual. This view of fairness, which we call fairness-as-sameness, emphasizes the importance of treating everyone the same, giving everyone the same formal opportunity to enter the competition for a position, and evaluating each person's results the same way. If everyone takes the same test, and every applicant's test is evaluated in the same manner, then the test is fair. Because affirmative action evaluates some people's test results differently, it is unfair.

This notion of fairness also implicitly involves a concept of notice and detrimental reliance. It is only fair that the rules governing selection remain constant throughout the process. If an employer or school conveys the expectation that it will select candidates based on test results and scores, and it then departs from those standards to take race or gender into account, it is breaching an implicit agreement to use particular criteria for hiring or admissions. ⁴¹ A crucial premise of this fairness challenge to affirmative action is the assumption that tests afford equal opportunity to demonstrate individual merit, and therefore are not biased. The presumption is that one-size-fits-all selection produces fair results.

The yardstick is the implicit, and sometimes explicit, metaphor for the version of merit and fairness that underlies the stock narrative. ⁴² The assumption is that institutions know what they are looking for (height), they know how to measure who has those characteristics (yards, meters), they can fairly replicate the measurement process (using the ruler), and they can rank people accordingly (by height). This approach to merit selection has been institutionalized in a manner that rests heavily on the use of standardized tests and other "objective" screening tools. Those who participate in selection, along with those who challenge how affirmative action operates, assume that these approaches to selection are generally valid and fair, and that they are justified as a means of pursuing merit. They assume that objective tests for particular attributes of merit, perhaps supplemented by subjective methods such as unstructured interviews and reference checks, are the state-of-the-art selection method, that they can be justified as predictive of performance, that they are the most efficient method of selection, and that no better alternative exists.

The stock narrative naturalizes the current yardstick model of merit. Thus, "more qualified" means that the applicant scored higher on standardized tests and interview ratings. Fairness means that applicants should be numerically ranked for selection using "objective," race- and gender-neutral selection criteria.

In fact, the dominance of standardized tests in selection is a relatively recent development. The civil rights revolution, and the introduction of affirmative action programs, occurred at the same time that society was formalizing a "meritocracy" based on education and standardized testing.⁴³ The construction of this meritocracy was part of an **[*966]** overall rationalization and formalization of the selection process,⁴⁴ in part as a response to the simultaneous legalization of norms for workplace conduct.⁴⁵ This move to objective testing arguably minimized both arbitrariness and individualized bias. In fact, many defended objective testing on the grounds that it opened up opportunity to people of all socio-economic backgrounds.⁴⁶ It eliminated the class-linked prerequisites to work and education that governed in the pre-standardized testing era, and instituted a system that presumably offered everyone a fair, unbiased, and equivalent chance to compete for educational and employment opportunities.⁴⁷

Yet, for blacks and other people of color, "rationality" was introduced into an environment that was not benign. When the civil service and educational institutions moved to standardize admission criteria, most jobs and institutions of higher learning were still segregated.⁴⁸ **[*967]** Both in the educational context and the workplace, opportunities had been virtually closed to blacks, except for those institutions made up almost entirely of blacks. By itself, rationality would not cure the hostility and exclusion, which were justified by some as a "rational" response to black inferiority.⁴⁹

Thus, affirmative action as a remedial response emerged when many of society's norms were being challenged. On the one hand, merit was increasingly judged on a single or dominant criterion of performance (the ability to get good grades or perform well on tests that are designed to assess general intelligence or inherent ability).⁵⁰ On the other hand, that single criterion of performance was exactly the area in which blacks had been made most vulnerable, factually, legally, and mythologically. As a factual and legal matter, blacks' educational opportunities had been severely limited. As the subject of political and pseudo-scientific mythmakers, blacks were pervasively stereotyped as possessing less general intelligence or inherent ability.⁵¹ Thus, the push for greater rationality in the workplace and institutions of higher learning must be juxtaposed against efforts by blacks and other people of color to challenge both the racially discriminatory allocation of benefits and the racial stereotypes that reinforced that allocation of benefits.⁵² **[*968]**

Consequently, we will examine the story of merit and unfairness from several perspectives. In Part II.A, we scrutinize the assumption that conventional approaches to selection in fact distinguish those who can perform well from those who cannot. In Part II.B, we examine the fairness of the so-called meritocracy, showing that existing selection processes undervalue many important skills and capacities, give undue advantages to those from higher socio-economic backgrounds, and exclude women and people of color at a disproportionate rate.

Unpacking Merit, Fairness, and the "Testocracy"

The stock narrative proceeds from and depends on the premise that the selection criteria and processes used to rank applicants for jobs and admission to schools are basically fair and valid. Yet, a substantial body of literature fundamentally challenges this basic premise. Even accepting the definitions of merit and fairness employed in the stock narrative, current approaches to selection are extremely limited in their predictive capability. If we apply a more comprehensive conception of fairness, the existing "meritocracy" fails even more miserably as a method of selection.

We argue that the "meritocracy" is neither fair nor democratic, neither genuinely predictive nor functionally meritocratic. Not everyone is being given an equal opportunity to compete. Not everyone who could do the job, or who could bring new insights into how to do the job even better, is being given an opportunity to perform or succeed.

Instead, a "testocracy" masquerades as a meritocracy. By testocracy we refer to test-centered efforts to score applicants, rank them comparatively, and then predict their future performance. Although subjective and idiosyncratic measures are often used as part of the assessment process, such criteria frequently supplement the basic approach of seeking to predict future open-ended performance through static, closed-book, timed paper-and-pencil assessments of past ability. These approaches to selection are neither fair nor functional. The yardstick metaphor simply does not hold up under scrutiny.

A. Merit and the Fiction of Functionality

This Section scrutinizes the claim that the conventional "meritocracy" functions to identify those who can best perform in the positions under consideration. For purposes of our argument, we accept the idea, without question, that functional capacity to perform, or functional merit, is a legitimate consideration in distributing jobs and educational opportunities.⁵³ We also assess merit as a functional concept related to the capacity to perform effectively, rather than as a concept of desert or societal values. The concept of merit as desert or earned recognition collapses either into a concept of formal fairness or of functional merit.⁵⁴ The concept of merit as societal or institutional values is harder to defend against a claim of exclusion, unless it has a functional connection to the position under consideration. Merit as a functional value offers the strongest and most widely embraced justification for embracing standards that exclude members of particular disempowered groups.

The stock affirmative action narratives equate merit with performance on standardized tests. These paper-and-pencil tests do not fulfill their stated function. They do not reliably identify those applicants who will succeed in college or later in life, nor do they consistently predict those who are most likely to perform well in the jobs they will occupy. Used alone or in combination with informal networking and subjective assessment, timed paper-and-pencil tests screen out applicants who could nevertheless do the job.

To fulfill their stated function, testers must be able to identify and measure

successful performance in the job or at school.⁵⁵ However, in both contexts, testers have failed to develop meaningful measures of **[*970]** successful performance. In the employment area, they have not attempted to correlate test performance with worker productivity or even with pay.⁵⁶ Instead, they rely on correlations between test performance and job sample tests or supervisor evaluations. Job test samples do not themselves correlate with productivity. Supervisor ratings are notoriously unreliable measures and have been shown to be biased in ways that correlate with race and gender.⁵⁷

In the educational context, testers attempt to correlate standardized tests with first-year performance in college or post-graduate education.⁵⁸ But this measure does not reflect successful overall academic achievement or performance in the areas valued by the educational institution.⁵⁹ Neither of the major testing services even attempts "to estimate the effect on college completion of different admissions policies, nor have they encouraged colleges to do so themselves."⁶⁰ Moreover, the data does not even consider the relationship between standardized test scores and a truly functional baseline--post-graduate "success" in life.⁶¹ Education and employment tests do not predict whether students will graduate or employees will succeed in their jobs.

Even if we accept the inadequate definitions of success used in conjunction with standardized tests, research shows a tenuous connection between test scores and successful performance. Recent studies show that the measured relationship between a test and predicted job performance, referred to as the correlation coefficient, is weak.⁶² The best employment tests have correlations of approximately .3.⁶³ Using the widely accepted statistical methodology for determining how much explanatory information a test provides,⁶⁴ a correlation of .3 means that "the test explains only 9% of the variation in predicted performance. **[*971]** In other words, the test leaves unexplained 91% of the variance reflected in the performance measure."⁶⁵

Validity studies of aptitude tests used to predict performance as measured only by first-year grades show correlations similar to those in the employment context. A recent study of the correlation of SAT scores with freshman grades showed correlations ranging from .32 to .36.⁶⁶ As David Owens notes, the correlation between "SAT scores and college grades ... is lower than the correlation between weight and height; in other words, you would have a better chance of predicting a person's height by looking only at his weight than you would of predicting his freshman grades by looking only at his SAT scores."⁶⁷ A recent study of the University of Pennsylvania Law School found that LSAT scores were weak predictors of performance in law school. LSAT explained 21% of the differences in third-year grades. For first- and second-year students, it explained even less: 14% and 15% respectively.⁶⁸ A study of the Texas Index at issue in the Hopwood case also **[*972]** found weak correlations between LSAT scores and first-year law school performance.⁶⁹ For black students, the relatively weak correlations between Index scores and first-year grades were achieved only "if lower undergraduate grade-point averages are made to predict higher first-year averages in the Law School."⁷⁰

Testing experts and decision makers acknowledge that the correlation between test scores and first year college performance for those who actually matriculate is relatively low.⁷¹ However, they point out that the group of people who are accepted by and attend a particular institution come from a

narrow subset of the total applicant pool.⁷² To deal with this problem,⁷³ experts have developed a formula that theoretically **[*973]** enables a projection from the performance of the known group of applicants to the performance of those who were not admitted and did not enroll.⁷⁴ This mathematical correction increases the correlation between the test and performance in the position.⁷⁵

However, the empirical basis for this mathematical correction has been seriously questioned.⁷⁶ The formula assumes continuity and similarity among applicants so that one can simply project out in a linear fashion from the performance of those who complete their freshman year.⁷⁷ But in many situations, we cannot know, for example, whether the pattern of performance of those who were admitted from the 90th percentile of test-takers would resemble the pattern of those who were not admitted from the 70th or 80th percentile of test-takers.⁷⁸ Some tests predict high test-scorers' performance better than that of lower test-scorers; others offer reliable information only about clear failures.⁷⁹ Yet, the mathematical adjustment assumes that the tests are equally predictive at different levels of performance on the test. Finally, this mathematical formula fails to take into account the many variables other than test scores and grade point average that might influence acceptance, enrollment, and completion of freshman year.⁸⁰ For these reasons, in some contexts experts have taken the position that "the **[*974]** conservative response is to apply no correction for restriction of range."⁸¹

Indeed, empirical and statistical evidence suggests that many of those who are excluded based on test results could perform comparably to those admitted. Many tests exclude applicants who could in fact perform successfully. A vivid example arose from an error in the scoring of the 1976 version of the Armed Services Vocational Aptitude Battery. A calibration error resulted in the admission to the military of over 300,000 recruits who actually failed the screening test used by the armed services.⁸² Studies examining the subsequent performance of those "potentially ineligible" ("PIs") found that performance differentials were "not large and in several cases the PIs performed as well as or even better than the controls."⁸³ The PIs "completed training; their attrition rates weren't unusually high; they were promoted at rates only slightly lower than their higher-scoring peers; and they reenlisted."⁸⁴

It is widely recognized that high school grades are more predictive of college freshman-year grades than the SAT.⁸⁵ Perhaps even more significant is the extremely small increase in predictiveness gained by using the SAT in conjunction with high school grades. "Colleges now make only trivial improvements in their ability to forecast college completion correctly when they use the SAT and rank together instead of using rank alone."⁸⁶ The incremental value of the SAT in predicting **[*975]** bachelor's degree attainment is even smaller.⁸⁷ Similar questions have been raised about whether cognitive tests used in the employment context have greater predictive value than alternative screening devices "such as education, job experience, peer evaluations, and interviews."⁸⁸

It is also difficult to justify the use of rankings to distinguish at the margins among people whose performance falls within a relatively narrow band. Our yardsticks of merit can be used to differentiate yards, perhaps, but not inches

or half-inches. The statistical concept of the standard error of measurement suggests that any particular test score is only indicative of an individual's "scoring range."⁸⁹ Although tests may offer useful information about people at the very top and very bottom of the pool, they cannot reliably differentiate among candidates at the margins or in narrow bands of test scores.⁹⁰ Yet many "reverse discrimination" cases involve candidates whose scores fall within relatively narrow bands that cannot reliably be distinguished.⁹¹ **[*976]**

An additional problem in establishing correlations between test scores and performance stems from the possibility that some people who may perform well in an educational or work environment perform poorly under the unique circumstances of most testing conditions.⁹² Moreover, most test instruments measure a wide range of skills and abilities through the narrow lens of the linguistic and logical-mathematical domains; if test-takers "are not strong in those two areas, their abilities in other areas may be obscured."⁹³ There is developing evidence that attributes that cannot easily be measured through standardized paper-and-pencil tests, such as discipline, emotional intelligence, commitment, drive to succeed, and reliability, may be more important to successful work or school performance than marginally better performance on tests of general intelligence or analytical ability.⁹⁴ Emotional qualities, what Daniel Goleman calls "emotional intelligence," may be just as important a predictor of academic success as test-taking ability.⁹⁵ Standardized tests "do not measure motivation, perseverance or teamwork skills."⁹⁶

A study of three classes of Harvard alumni over three decades, for example, found a high correlation between "success"--defined by income, community involvement, and professional satisfaction + and two criteria that might not ordinarily be associated with Harvard fresh- **[*977]** men: low SAT scores and a blue-collar background.⁹⁷ When asked what predicts life success, college admissions officers at elite universities report that, above a minimum level of competence, "initiative" or "hunger" are the best predictors.⁹⁸

Success may simply reflect a person's opportunities to learn a job or a skill, opportunities that are not measured by any paper-and-pencil test.⁹⁹ One-size-fits-all testing may also compromise the institution's capacity to search for what it really values in selection. Researchers are becoming increasingly aware that privileging the aspects of performance rewarded by standardized tests may well screen out the contributions of people who would bring important and different skills to the workplace or educational institution.¹⁰⁰ Finally, individual performance in both the workplace and educational environments is often enhanced when challenged by competing perspectives or when given the opportunity to develop in conjunction with the different approaches or skills of others.¹⁰¹

The problem of using standardized tests to predict performance is even more acute in the employment context. Standardized tests may reward qualities such as willingness to guess, conformity, and docility.¹⁰² If this is so, then test performance may not relate significantly to the capacity to function well in jobs that require creativity, judgment, and leadership. In a service economy, creativity and interpersonal skills are important, though hard to measure. In the stock scenario of civil service exams for police and fire departments, traits such as honesty, courage, and ability to manage anger are left out.¹⁰³ In other words, people who rely heavily on numbers to make employment decisions

"are being misled." ¹⁰⁴ [***978**]

Thus, if John Doe scored higher on the civil service exam, that does not necessarily mean that he would perform better as a police officer. Cheryl Hopwood's higher Texas Index score does not necessarily establish her superior qualifications to be a law student (or a lawyer). Yet, many employers and schools effectively use test scores to rank candidates or establish cut-off scores within relatively narrow bands of test performance. ¹⁰⁵ Colleges and professional schools often use cut-off scores or presumptions to create categories of students who will be automatically admitted or rejected. ¹⁰⁶ Small, statistically insignificant differences in test scores may well determine whether a candidate is admitted or rejected. ¹⁰⁷ Schools and employers use these tests even while they admit that the tests have very limited predictive value. Similarly, complainants in cases challenging affirmative action policies also rely on their test scores, even though those scores may not, in statistical terms, tell us very much about their likely performance in a job or in law school. ¹⁰⁸

Advocates of objective tests usually respond that, although these tests are limited, they are the best we can do. One of the most common and facially compelling arguments for retaining uniform aptitude tests, such as the SAT, focuses on their role as a "leveling agent":

In addition to its supplemental and incremental value, the SAT has a value of its own in confirming the grades from different schools. Since the high school record is a reflection of locally controlled curricula and local grading practices ... there are variations from school to school in the meaning of the [***979**] grades... The SAT, on the other hand, represents a standardized measure of the same mental tasks... ¹⁰⁹

This argument is often supplemented by concerns that eliminating the SAT would induce high schools to give more A's. ¹¹⁰ Thus, standardized tests are offered as a solution to the problem of grade inflation and nonuniformity.

However, this stated concern about the difficulty of interpreting local differences in grading practices does not account for the practices actually used to assess high school records. Colleges often improve their capacity to evaluate high school performance by relying on rank in class, rather than simple grade point average, and can demand that high schools supply grade distributions for their entire graduating classes. ¹¹¹ In addition, colleges and professional schools have relationships and track records with institutions that enable them to make informed judgments. ¹¹²

More importantly, the data undercuts the significance of the "leveling" argument as a basis for relying on standardized tests to improve selection decisions. For purposes of selection, grade inflation and nonuniformity in grading practices only matter if they undercut the value of high school grades in predicting performance in college. The evidence suggests exactly the opposite. Whatever the variability in local grading practices, high school grades have consistently proven more predictive of freshman grade point average than the SAT. ¹¹³ This has remained true for both selective and non-selective colleges. ¹¹⁴ Indeed, studies show that during the period of 1981-88, the

predictive value of high school grades has increased for more selective institutions, while the predictive value of SAT scores has decreased slightly.¹¹⁵ During a period in which high schools may have experienced grade inflation, there was no detectable decline in the predictive value of high school grades.¹¹⁶ During this same period, the SAT did not appreciably improve colleges' capacity to select those most likely to perform well in their freshman year.¹¹⁷

Another important justification for relying on standardized tests stems from their cost-effectiveness. Standardized tests can be administered to huge numbers of applicants at relatively low cost. This view of "cost-effectiveness" focuses on short-term expenses of selection. It fails to take into account the costs to institutions of using selection criteria that do not predict successful performers.¹¹⁸ It denies institutions that use tests to screen applicants the capacity to admit potential leaders from unconventional backgrounds, whose skills are not easily quantifiable.¹¹⁹ It also fails to consider the full range of costs incurred in developing and relying predominantly on standardized tests to predict performance. These costs include the social consequences of excluding those who have the capacity to succeed but are "stigmatized" by their weak performance on the tests.¹²⁰ They also include the costs of failing to explore more innovative and potentially productive ways of selecting candidates. As we show in the next Section, it also tends to give the wealthy a thumb on the scale of merit.

B. Standardized Testing and the Fiction of Fairness

The previous Section challenges the assumption, implicit in the stock affirmative action narratives, that those who score higher on standardized tests will function better in the positions they seek. This Section takes on the second implicit assumption of those narratives: that the conventional methods of selecting candidates for high-stakes positions are fair. To assess the implicit claim that existing selection criteria are fair, it is crucial to broaden the conception of fairness that frames the analysis.¹²¹ The stock affirmative action narrative implicitly embraces process definitions of fairness: do applicants receive the same treatment in the evaluation process? Are employers adhering to the stated standards for everyone? This concept of fairness is misleading. It presumes a level playing field--that if everyone plays by the same rules, the game does not favor or disadvantage anyone.

There is, however, a conception of fairness, which we call "fairness as equal access and opportunity," that summons the substantive dimension beneath formal sameness. This more substantive conception challenges the assumption that in all situations sameness equals fairness. It focuses on providing members of various races and genders equivalent opportunities to demonstrate their capacities. It focuses attention on the fact that formal sameness camouflages actual difference.¹²² And it focuses attention on differential access and the exclusionary bias built into the screening devices used to allocate positions.

This conception of fairness requires that the standards governing the process not arbitrarily advantage members of one group over another. If different approaches can be used to accomplish an underlying goal, and these differing approaches correlate with race or gender, then an employer may not fairly

insist on only one of these approaches. ¹²³ **[*982]** Alternatively stated, fairness must rest on genuine equality of opportunity. ¹²⁴ Fairness means that "everyone should be given an equal opportunity to achieve all that their abilities allow." ¹²⁵ It is not "fair" in this sense to use entry-level credentials that appear to treat everyone the same, but in effect deny women and people of color a genuine opportunity to demonstrate their capacities. ¹²⁶

In fact, the "testocracy" does not provide a fair playing field for candidates. First, many standardized tests are substantively unfair because they assume that there is a single, uniform way to complete the job, and then tests applicants solely upon criteria consistent with this uniform style. In this way, the testing process entrenches the status-quo mode of production, excluding those individuals who may perform the job just as effectively through different approaches. Second, conventional selection methods advantage candidates from higher socio-economic backgrounds and disproportionately screen out women and people of color, as well as those in lower-income brackets. When combined with other unstructured screening practices, such as personal connections and alumni preferences, standardized testing creates an arbitrary barrier for many otherwise-qualified candidates. **[*983]**

1. The Underinclusiveness of One-Size-Fits-All Tests: Sameness Is Not Fairness

Standardized tests adopt a one-size-fits-all approach to measuring successful performance. In addition to ignoring many of the abilities and skills that are crucial to successful performance, ¹²⁷ this insistence on narrow, uniform criteria of success fails to take account of the variety of ways in which successful performance on the job can be achieved. There may be a range of styles and approaches to doing a job, each of which may be effective in some circumstances. Indeed, diversity introduces a variety of job approaches that can complement one another and offer new and potentially more effective styles and strategies.

For example, in many police departments, strength, aggressiveness, and speed are the predominant criteria of selection for police officers. ¹²⁸ These characteristics relate to a particular mode and concept of policing focusing on "command presence" and control through authority and force. ¹²⁹ If the issue is quick reaction time and physical prowess, some of these qualities, such as speed, may be important. But not everything a police officer does requires quick reaction time. Indeed, in some situations, responding quickly gets police officers and whole departments in trouble. ¹³⁰ **[*984]**

This speed-and-strength standard normalizes a particular type of officer: tough, brawny, and macho. ¹³¹ However, studies have begun to show that other modes of policing, such as dispute resolution, persuasion, counseling, and promoting community involvement, are also critical and sometimes superior approaches to policing. ¹³² One study of the Los Angeles police department, conducted in the wake of the Rodney King trials, recommended increasing the representation of women on the police force as a critical component of a strategy to reduce the level of police brutality and improve community relations. The study found that women often display a more interactive and engaged approach to policing. ¹³³