Psychology, Public Policy, and Law
2000, Vol. 6, No. 1, 129–143

1076-8971/00/$5.00 DOI: 10.1037//1076-8971.6.1.129

SKILLS GAPS, NOT TESTS, MAKE RACIAL PROPORTIONALITY IMPOSSIBLE

Linda S. Gottfredson
University of Delaware

Recent bans on racial preferences in some states have increased the political pressure to abandon cognitive testing to raise rates of minority hiring and college admissions. However, general skills gaps, not tests, are the major remaining impediment to racial parity in outcomes. Data on the size, stability, and functional importance of racial gaps in cognitive skills show the futility of seeking racial parity without first closing skills gaps. Lowering mental standards for Blacks alone (double standards) or for all races (low common standards) far enough to produce Black–White parity in educational and occupational outcomes would lead to performance problems so conspicuous and so severe as to force a reversal of that effort to achieve proportionality. There is widespread reluctance to acknowledge that racial skills gaps are a serious national problem, but confronting them energetically offers the most hope.

Tests of cognitive ability and achievement perform a useful social and economic function; namely, they rank individuals according to skills that enhance performance in school and work. Test use for admitting students and selecting workers waxes and wanes, however, as legal and political impediments to that use lose or gain force. The recent upsurge in antitest sentiment results from new bans on racial preferences in college admissions in California and Texas. One headline succinctly described the new pressure: "As states end racial preferences, pressure rises to drop SAT [Scholastic Aptitude Test] to maintain minority enrollment" (Kronholz, 1998).

The state of Texas has done just that. After a federal appeals court ruled that racial preferences in admissions are illegal, the state legislature passed a law that the top 10% of graduating seniors from all Texas high schools may enter any state university, regardless of their SAT scores. Not surprisingly, some high schools are now reporting 25% of their students in the top 10% of their class (Kronholz, 1998).

Rosen (1998) describes how written tests are being done away with across California, in civil service hiring as well as in college admissions, since Proposition 209's ban on racial preferences in the public sector took effect in California. The University of California is considering dropping the SAT. The University of California, Berkeley, Law School is considering dropping or deemphasizing the Law School Admission Test (LSAT): "Because the gap in scores is so wide, it's simply not possible to admit many African-Americans without taking race into account if the LSAT remains an important factor in the admissions process" (Rosen, 1998, p. 62). The Law School has already voted to stop weighting the grade point averages of applicants to reflect the quality of their undergraduate

Correspondence concerning this article should be addressed to Linda S. Gottfredson, School of Education, University of Delaware, Newark, Delaware 19716. Electronic mail may be sent to gottfred@udel.edu.

schools. "A 4.0 from California State is now worth as much as a 4.0 from Harvard" (Rosen, 1998, p. 62).

The pressing question for many has become, "What definition of merit will prevent the resegregation of top-ranked schools in a post-affirmative-action world?" (Rosen, 1998). The answer across much of California seems to be that "merit" should consist less of the accomplishments that White and Asian applicants exhibit more often than do Blacks and Hispanics and more of whatever attributes disproportionately characterize the latter. So, for example, colleges are giving less and less credit for how good an applicant's relevant aptitudes (test scores) and achievements (grades) are but more credit for how bad an applicant's environment has been (poverty, broken family, and so forth). To increase minority admissions substantially, however, common entrance standards must be lowered quite substantially. As Rosen (1998) notes, such a move threatens "the destruction of the great public universities." Apparently assuming that racial proportionality is the first priority of public policy, he therefore urges a return to racial preferences as a form of "damage control."

This battle over the use of cognitive tests in college admissions is but the latest highly public display of a policy already at work less visibly in employment settings: If one cannot lower standards just for lower-scoring minority groups, then one should experiment with changing standards for everyone. Like colleges, many employers had long been covertly using racial preferences to increase the proportion of minorities they selected. They did so for the same reason that most colleges still use both the SAT and racial preferences when admitting students. Cognitive tests measure thinking skills that are known to be crucial in education and work but that Black and (to a lesser extent) Hispanic applicants tend to possess at relatively lower levels. When Congress banned the practice of race-norming employment tests (ranking test takers' scores separately by race), that 1991 ban precipitated among employers the same panicked scramble for alternatives now occurring in university systems that may no longer use racial preferences in admissions.

On the positive side, there arose in employment testing a concerted new effort to identify and measure relevant noncognitive traits. Assessing both cognitive and noncognitive skills can enhance the predictive validity of employment selection batteries at the same time that it increases minority representation (because there are few racial differences in relevant noncognitive traits). The increase typically falls far short of proportionality, however (Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997). Not surprisingly, then, most new alterations in test use for increasing minority hiring have involved degrading selection for mental competence in any race by reducing either test reliability or test validity (for example, by "banding" cognitive test scores or giving little weight to mental ability or job knowledge relative to personality traits).

When the bottom-line requirement becomes racial proportionality in outcomes despite large gaps in relevant skills, something that the U.S. Department of Justice now seems to insist on in police and fire departments nationwide, that requirement entails hiring virtually without regard to mental capability, no matter how important it may be for good job performance (Gottfredson, 1996; Schmidt, 1996). This stripping away of meaningful mental standards is generally defended in the name of improving selection procedures, for example, by rejecting "mean-

ingless" distinctions on "narrow" (objectively measured and validated) cognitive traits (reasoning, analytical ability, problem solving, and general intelligence) in favor of assessing the "whole person" in a more "flexible" or "innovative" (subjective, unvalidated, and race-driven) manner. Ignoring mental competence when it is important, however, guarantees that the organization will soon taste the fruits of incompetence. Open admissions destroyed the reputation of the once great City University of New York. In like fashion, public safety deteriorates quickly in police jurisdictions that abandon mental standards in hiring, as Washington, DC, discovered in the 1980s (Zelnick, 1996, p. 107–118).

Personnel psychologists have long shown that low common standards can be expected to devastate productivity immediately, whereas racial preferences depress workforce performance considerably less in the short term. These findings have led some to suggest that double standards are to be preferred over low standards. What they ignore, however, is not only that racial preferences are constitutionally suspect and socially divisive but also that they are just a slower and more corrosive way of ending up with low common standards (e.g., see Gordon, 1988). Moreover, neither strategy for achieving racial proportionality can ever produce an enduring racial equality, because, as I shall show, neither strategy circumvents the real-world consequences of the skills gaps that they ignore. Those real-world consequences are typically large and visible enough to stall minority hiring and college admissions far short of proportionality. The real choice in seeking racial progress thus concerns not how to ignore racial gaps in mental proficiency but how to confront them.

Skills Deficits as a Barrier to Racial Proportionality

I focus here on the Black–White gap in cognitive skills because it is the largest and also the major impetus behind racial preferences. An initial intuitive sense of its magnitude can be gleaned from the fact that Black youth of college age tend to lag 4 years behind White youth in basic academic skills. This lag means that Black 17-year-olds possess skills in reading, writing, mathematics, and science that are comparable to those of white 13-year-olds (Smith et al., 1995, pp. 54, 56, 58, 60) and that Black college graduates have literacy skills comparable to those of age-matched White persons possessing only a high school diploma (Kirsch, Jungeblut, Jenkins, & Kolstad, 1993, p. 127).

Black-White Gaps in Functional Literacy

Table 1 shows the percentages of White and Black adults who function at each of five literacy levels assessed in the U.S. Department of Education's National Adult Literacy Survey (NALS; Kirsch et al., 1993). The sample items illustrate the general level of information processing at which individuals at that literacy level routinely function. All the literacy items comprising the three scales reflect everyday demands for functional competence, for example, in dealing with banks, social welfare agencies, and restaurants and understanding public issues, events of the day, and one's personal options for services and benefits. As an illustration, individuals at Level 1 have about an 80% probability of correctly performing tasks that are no more complex than locating the time of a meeting

Table 1
Percentages of Whites and Blacks ≥16 Years Old at Five Levels of Proficiency on Three National Adult Literacy Survey Scales

Proficiency level (scores)	Approximate grade level[a]	Percentage with the indicated score on the following scale: Prose[b] White	Prose[b] Black	Document[b] White	Document[b] Black	Quantitative[b] White	Quantitative[b] Black	Sample item (scale value)[c]
1 (≤225)	2.5	14	38	16	43	14	46	Total a bank deposit entry (190) Locate one piece information in sports article (210)
2 (226–275)	7.2	25	37	27	36	24	34	Locate intersection on street map (230) Locate two pieces information in sports article (250)
3 (276–325)	12	36	21	34	18	35	17	Write brief letter explaining credit card bill error (288) Calculate miles per gallon from mileage record chart (321)
4 (326–375)	16	21	4	19	3	21	3	Determine correct change from information in menu (331) Contrast views of two editorials on fuel-efficient cars (359)
5 (376–500)	n.a.	4	0	3	0	5	0	Determine shipping and total costs on order form (382) Interpret brief phrase from long news article (423)

[a] Levels 1 to 4 of the Adult Prose Scale correspond, respectively, to the rudimentary (150) plus basic (200), intermediate (250), adept (300), and advanced (350) National Assessment of Educational Progress reading levels reported for school-age children. Carroll (1987, p. 425) has estimated that they correspond, respectively, to grade equivalents of 1.5 (rudimentary) and 3.6 (basic), 7.2 (intermediate), end of 12 (adept), and 16 (advanced). n.a. = not applicable.
[b] Data are from Kirsch et al., 1993, Table 1.1A to 1.1C.
[c] Items are from Kirsch et al., 1993, Figure 1.

specified on a form or one particular piece of information in a sports article. Items at successively higher literacy levels require processing increasingly complex information. As shown in Table 1, one of seven White adults but two of five Black adults routinely function at Level 1, the lowest of the five. In contrast, about 1 of 4 Whites but only 1 of 30 Blacks functions at Level 4 or above, which requires the ability to draw inferences and integrate information from long or complex materials.

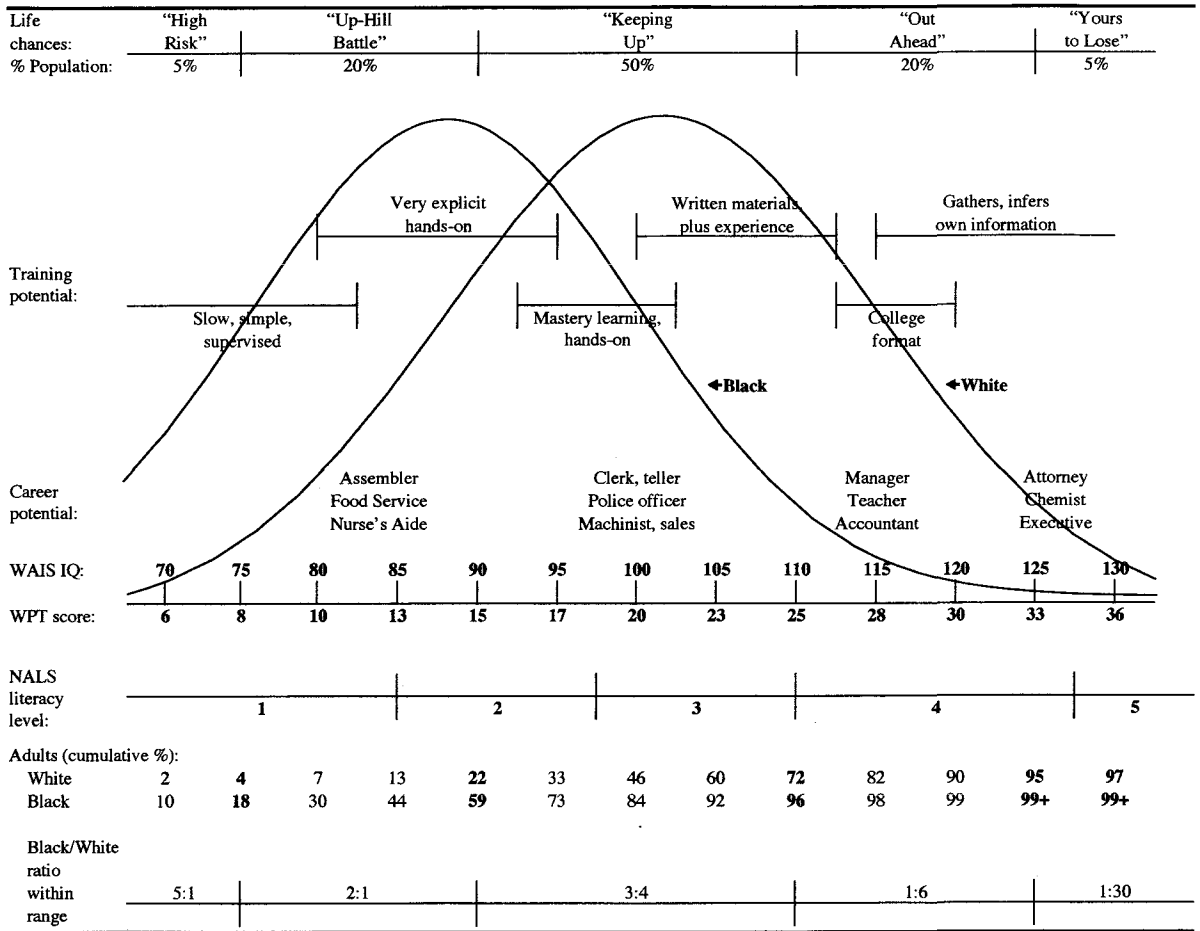
These are highly consequential racial disparities. Individuals at Level 1 or 2 "are not likely to be able to perform the range of complex literacy tasks that the National Education Goals Panel considers important for competing successfully in a global economy and exercising fully the rights and responsibilities of citizenship" (Baldwin, Kirsch, Rock, & Yamamoto, 1995, p. 16). This population would include two fifths of White adults but three quarters of Black adults. For racial proportionality to be achieved in the presence of these skills disparities, Blacks would routinely have to work or study beside Whites with functional literacy skills one level higher, and these differences in proficiency would quickly become obvious to coworkers and supervisors. As an illustration, adults performing at a score of 200 (in Level 1) have an 80% chance of performing Level 1 tasks correctly but only about a 30% chance of performing correctly the average task at Level 2 (Kirsch et al., 1993, pp. 101–102). Workers with Level 1 proficiencies might generate more problems than they solved in Level 2 jobs or require more supervision than their labor was worth, thus creating a net loss for employers even if the workers were paid no salary.

Individual and group differences in basic skills develop slowly over the elementary and secondary school years (Carroll, 1987). Neither colleges nor employers can be expected to narrow those differences quickly, if at all. Although many skills of the sort listed in Table 1 perhaps can be taught to many people who routinely function below a given level, those items represent only specific manifestations of highly general cognitive processing capabilities that are at best exceedingly difficult to improve across the board. For example, persons at Level 1 might be taught how to correctly perform certain Level 2 skills (such as locating a given intersection on a street map), but this process would not improve their performance of the myriad other individual Level 2 skills that they were not specifically taught (such as interpreting instructions from an appliance warranty) but at which persons at Level 2 would already be proficient because of their better information-processing capabilities. The more specific the skills, the more likely that they can be improved, but there is no evidence that individual differences in broad ranges of such skills can be substantially eliminated.

Intelligence, Employability, and Trainability

As already indicated, the NALS scales actually measure general information-processing skills. Evidence suggests, in fact, that they measure *g*, or phenotypic intelligence (Gottfredson, 1997b). Much is known about the general trainability and employability of individuals along different ranges of the IQ continuum. Figure 1 summarizes such data, showing also the IQ ranges to which the five NALS levels correspond.

Figure 1 indicates, for example, that clerical work and crafts work draw


Life chances:
"High Risk"
"Up-Hill Battle"
"Keeping Up"
"Out Ahead"
"Yours to Lose"
% Population:
5%
20%
50%
20%
5%
Training potential:
Slow, simple, supervised
Very explicit hands-on
Mastery learning, hands-on
Written materials plus experience
College format
Gathers, infers own information
←Black
←White
Career potential:
Assembler
Food Service
Nurse's Aide
Clerk, teller
Police officer
Machinist, sales
Manager
Teacher
Accountant
Attorney
Chemist
Executive
WAIS IQ:
70 75 80 85 90 95 100 105 110 115 120 125 130
WPT score:
6 8 10 13 15 17 20 23 25 28 30 33 36
NALS literacy level:
1
2
3
4
5
Adults (cumulative %):
White 2 4 7 13 22 33 46 60 72 82 90 95 97
Black 10 18 30 44 59 73 84 92 96 98 99 99+ 99+
Black/White ratio within range
5:1
2:1
3:4
1:6
1:30

most of their workers from the middle of the IQ bell curve, whereas semiskilled and professional occupations tend to recruit, respectively, from somewhat below-average and considerably above-average segments of the IQ distribution. This recruitment pattern is consistent with data indicating that the ability of mental tests to predict job performance rises with job level (see Hunter & Schmidt, 1996, for a review of these and other relevant data). That is, the same 10-point IQ difference between workers will have a larger average effect on performance in more complex jobs, and bright workers will thus be even greater assets in more cognitively demanding occupations. No occupation routinely recruits workers from below IQ 80 (the 10th percentile), nor is the military allowed to do so, except in times of war. Employability is clearly low below IQ 80 but rises steadily with IQ.

Trainability likewise rises with IQ level. Training must be highly concrete, specific, and tightly supervised at low IQ levels but can be increasingly abstract and self-directed at higher IQ levels. Very bright people are self-training and learn readily from experience, which is essential in nonroutine, critical jobs at the top of the occupational hierarchy. Current military enlistment standards are set above the 16th percentile (about IQ 85). Past research on recruits between the 10th and 16th percentiles of general ability showed that they were very difficult and costly to train, could not learn certain specialities, and performed at a lower average level once on a job (Laurence & Ramsberger, 1991). Analysts who nonetheless favor the enlistment of such recruits (calling them "cast-off youth") are particularly explicit about "how difficult it is for the [military training] schools to train personnel of all aptitudes when slow learners may require two to five times more instructional time than more able learners" and about how success in training low-aptitude men "in some cases . . . may be achieved if the training content is limited strictly to that which is relevant to a specific job, and no attempt is made to supply any underlying theory or more general instruction which might be useful to trainees of higher aptitude in fitting them for rapid advancement to positions of greater responsibility" (Sticht, Armstrong, Hickey, & Caylor, 1987, pp. 91, 94).

As depicted in Figure 1, people with IQs below 75 tend to be at "high risk" for failure in modern, industrialized societies no matter what their race is. Life is an "uphill battle" today for most people in the IQ range of 75 to 90, regardless of their social origins. In contrast, success is "yours to lose" for people in the IQ range above 125. As the information age advances, highly able people are increasingly competitive and cognitive deficits are increasingly a handicap (Hunt,

---

Figure 1. Typical levels of adult literacy, employability, and trainability at different ranges of the IQ bell curve. Cumulative percentages for adults were based on mean Wechsler Adult Intelligence Scale (WAIS) IQs of 101.4 for Whites and 86.9 for Blacks and *SD*s of 14.7 for Whites and 13.0 for Blacks (Reynolds, Chastain, Kaufman, & McLean, 1987, p. 330). Percentiles for IQ scores were estimated by use of cumulative normal probability tables. Black/White ratios were calculated before percentiles were rounded. WPT = Wonderlic Personnel Test; NALS = National Adult Literacy Survey. Reprinted from "Why *g* Matters: The Complexity of Everyday Life," by L. S. Gottfredson, 1997, *Intelligence, 24,* Figure 3, p. 117. Copyright 1997 by Elsevier Science. Reprinted with permission.

1995). This notion should not be surprising because *g* is essentially the ability to process complex information quickly and efficiently. It represents differences in what are often referred to as higher-order thinking skills.

Racial Disproportions in IQ

The foregoing patterns of employability and trainability help to illustrate the implications of the one-standard-deviation (15-point) difference between American Whites and Blacks on valid, unbiased tests of (phenotypic) intelligence. The Hispanic–White difference is about half that, although it varies by subgroup (Cuban, Puerto Rican, and so forth). Asian Americans score at least as well as Whites on tests of intelligence, although they too differ by specific ethnic group (as do Whites).

As shown in Figure 1, a one-standard-deviation difference in IQ represents a substantial average difference in trainability and employability. There are successive increments in average IQ of roughly one standard deviation between assemblers, bank tellers, store managers, and attorneys, for example. Whereas the White average is about 100, which is typical of workers in jobs of moderate prestige and complexity (clerical, protective service, and crafts workers), the Black IQ bell curve is centered at about 85, a score that is typical of semiskilled workers.

Current minimum military enlistment standards are set above IQ 85, thus excluding at least half of Blacks. Indeed, a 1980 U.S. Department of Defense study showed that whereas 71 to 89% of White men 18 to 23 years old were eligible for enlistment on the basis of aptitude and education (the Air Force being the most stringent and the Army the least), the percentages were only 21 to 41% for Black men and 38 to 53% for Hispanic men (Eitelberg, 1988, p. 99). Figure 1 illustrates why the disparate impact of valid, race-neutral hiring becomes even more acute in higher levels of education and employment. The proportion of Blacks who are available at successively higher levels of IQ drops dramatically relative to that of Whites. The Black–White ratio changes from 5:1 for IQs below 75 to 1:30 for IQs above 125.

Noncognitive traits such as conscientiousness are important for good job performance (Hunter & Schmidt, 1996). Even when such noncognitive traits might substitute for deficits in cognitive capability, however, they would have to be higher on average among Blacks and Hispanics to compensate for these groups' lower average cognitive skills. Research reveals no such compensatory noncognitive advantage for performing cognitively demanding tasks.

Side Effects of Seeking Racial Proportionality Despite Continuing Skills Gaps

Achieving proportionality in middle and high levels of education and employment without first reducing the Black–White skills gap can be expected to produce serious side effects. Specifically, if race preferences were used in an attempt to achieve proportionality, entrance standards would systematically have to be set one standard deviation lower for Blacks than for Whites. Figure 1 shows four standard deviations in IQ and four general tiers of work beginning at IQ 80. Proportionality would mean that Blacks recruited to any particular tier of work would tend to have mental skills comparable to those of White workers in the next

lower tier of work. For example, Blacks selected into the occupations of attorney, chemist, and executive would, on average, have the mental ability typical of Whites employed as managers, teachers, and accountants. In like manner, Blacks employed in the second highest tier would have mental capabilities more like those of White clerks, police officers, machinists, and other skilled workers. Blacks in the third highest tier would be comparable in mental ability to White food service workers, assemblers, and other semiskilled workers. That lowest tier of work would have to utilize many Blacks with IQs in the retarded range (below IQ 75).

Two thirds of a job's incumbents typically fall within a range of 15 to 17 IQ points, or about one standard deviation in IQ; 95% generally fall within a range of 30 to 34 IQ points (Gottfredson, 1997b, p. 90). A one-standard-deviation preference would mean that the IQ of the average Black worker in a job would be one standard deviation below that of the average White worker in the job. The White average IQ would be at the midpoint of the aforementioned 30- to 34-IQ-point range for the job, meaning that the average IQ for Blacks would be at the bottom of that 95% range (or the 2.5th percentile of White incumbents). Minimally acceptable performance is generally defined somewhere between the 10th and 16th percentile of workers in a job, meaning that a clear majority of Blacks would perform below that minimally acceptable level in a system that recruited them under a one-standard-deviation preference. However, regardless of how Blacks were officially evaluated on the job, their high rates of failure relative to the performance of Whites would become obvious. "Black" and "failure" would largely coincide (for an example, see Gottfredson, 1994b, p. 961).

Moreover, it is not clear that the same training programs could be used for Blacks recruited from IQ levels averaging one standard deviation below those of Whites. For example, whereas most White recruits to first-tier jobs would be readily trained with highly abstract materials and much independent learning, Black recruits to that tier would tend to require more concrete materials and guided experience, as would Whites of the same average IQ level. In the lowest tier of jobs, Whites would generally succeed in learning the jobs with highly explicit formal training, but Blacks (who, again, would have IQs averaging one standard deviation lower) would more often require the intensive, one-on-one, step-by-step training that borderline mentally retarded individuals need to succeed in nonsheltered employment, as would Whites or any other group at the same lower IQ level. In short, it is inconceivable that the one-standard-deviation racial preference necessary for racial proportionality in hiring would be sustainable in many settings for very long. Even a one-half-standard deviation racial preference might be difficult to sustain, and it would produce hiring and admissions rates far from parity in the top two tiers (respectively, .22 and .35; Gottfredson, 1986, p. 400).

Seeking racial proportionality via race-neutral means instead would require essentially randomizing selection above some minimum IQ level. That level would have to be very low, as already noted. For example, lowering the implicit IQ threshold for graduate and professional training to IQ 100 from its current level of about IQ 115 and then selecting randomly above that point would reduce the disparate impact ratio from .1 (about 2% of all Blacks vs. 18% of all Whites above that level, as shown in Figure 1) to .3 (about 15% vs. 54%), which is still far from

parity (1.0). Nor is it close to meeting the Federal Government's "four-fifths" standard (.8), the level below which organizations become vulnerable to discrimination lawsuits. The four-fifths rule would be satisfied only by setting minimum standards at about IQ 77 or 78, which is near the threshold for borderline mental retardation and far below the military's current minimum enlistment standards. It would be tantamount to random selection from the general population for all colleges and occupations. Random selection on mental ability above even IQ 100 (the 50th percentile) would be unthinkable for many colleges and occupations and, if implemented, would dramatically alter their character. In addition, as noted, the disparate impact ratio in that scenario would rise only to about .3.

Drop in Standards Halted Before Producing Racial Proportionality

Racial proportionality clearly is not going to be achieved by jettisoning mental standards either for all races or for just the groups with disproportionate numbers of skills-deficient members. Most organizations that adopt such policies will stop or reverse direction as the real-world consequences of incompetence become increasingly visible and intolerable. Gordon (1988) refers to this effect as reaching the "point of organizational embarrassment." He illustrated the phenomenon with Harvard Medical School which, in the 1970s, began lowering various academic demands in medical training in order to reduce or disguise racial disparities in student performance, which in turn had resulted from its lowering of admissions standards for minority students. Standards reached the point of embarrassment when one student to whom Harvard had granted a medical degree failed the National Board Examination a fifth time, prompting one faculty member to make public the school's domino-like collapse of training standards.

Mental standards will tend to oscillate near the point of organizational embarrassment, as the organization first shrinks back from the mounting side effects of falling standards but then ventures forth once again toward its racial goal, to be repelled yet again by the untoward effects of that pursuit. Few standards will ever oscillate far enough downward for long enough to produce even a temporary racial parity in hiring and admissions, except in atypical circumstances. Having compromised its standards for some or all races but yet failed to achieve its racial aim, the organization will have angered both the supporters and the opponents of the race-driven compromise. If the organization institutes racial preferences in the pursuit of proportionality, those preferences will be large enough to infuriate Whites but yield results too small to satisfy lower-scoring minorities. If it lowers standards for everyone to meet its racial goal, the organization will sustain significant damage from inept members but, once again, will not produce the desired racial parity in representation. In short, organizations that attempt to defy the functional importance of cognitive skills in order to increase minority representation trap themselves in a perverse purgatory, where to expiate one mortal sin requires committing another.

Reluctance to Acknowledge Racial Gaps in Skills

Discussions of racial disparities in cognitive skill are often painful, especially for Blacks who have labored under the burden of perceived racial inferiority for so long. Accordingly, there is widespread reluctance to acknowledge the reality

and importance of these disparities. Many scholars who acknowledge them privately will not do so publicly (Snyderman & Rothman, 1988), and some even argue that the nation is better served by telling reassuring lies rather than unsettling truths about race (e.g., Glazer, 1994; Gottfredson, 1994a).

Most reluctance among scholars to take the reality seriously is expressed, however, by doggedly disputing the scientific evidence for it, no matter how strong. This practice usually takes the form of trying to discredit at least one of three crucial links in the argument that racial skills gaps pose serious sociopolitical problems, namely, that skills gaps are real, stubborn, and functionally important. The network of evidence supporting these conclusions is now truly massive (e.g., Brody, 1992; Jensen, 1998; Neisser et al., 1996; Wigdor & Garner, 1982a, 1982b; see also the bibliography in Gottfredson, 1997a), so many skeptics have switched from finding fault with the supporting evidence to mostly ignoring it and pointing instead to wisps of evidence that can be construed to contradict one or more of the three conclusions. Several examples for each will illustrate this resolute search for doubt.

Steele's (1997) work on stereotype threat is increasingly cited to dispute the claim that racial gaps in test scores represent meaningful differences in cognitive skills. Even if one accepts the claim, however, that Steele succeeded in inducing a score-depressing anxiety among Black college students, no one has demonstrated that such test anxiety accounts for any significant proportion of the large racial gaps in mental test scores revealed in countless studies in diverse settings. Moreover, if stereotype threat were actually a significant source of racial differences in test performance, one should find, among other things, that mental tests generally underestimate Blacks' later performance in school and work, that test results are sensitive to race and the emotional supportiveness of the tester but not to the mental complexity of the task, and that racial gaps in test scores rise and fall with changes in the racial climate. Accumulated research, however, reveals quite the opposite (e.g., Jensen, 1980, 1998). Moreover, the consistency and continuity of results over time and place suggest that the source of current racial gaps is a much less labile phenomenon than Steele has shown stereotype threat to be. Theories of performance anxiety can explain the phenomenon revealed by Steele's recent studies of test-taking behavior (Jensen, 1998, pp. 513–515), but his theory of stereotype threat cannot explain the vast pattern of evidence generated in diverse fields that has led virtually all experts to believe that racial skills gaps are real. It is the sometimes overdrawn conclusions from Steele's studies, not the aforementioned body of well-established evidence, that should immediately be called into question.

Similarly ambiguous bits of evidence have been offered to counter the amply grounded conclusion that individual and group differences in general mental ability are highly stable and have so far resisted intervention (e.g., Moffitt, Caspi, Harkness, & Silva, 1993; Spitz, 1986). Those challenges typically exploit or foster conceptual confusions, perhaps the most common one being to imply that if one sees the level of some trait change over time, then that trait must be changeable in the sense of being malleable. Thus, for example, some critics wrongly imply that the great growth in mental capability and concomitant changes in brain structure that all children experience indicate that differences in the developmental outcomes among them must be readily malleable. Children grow considerably

in height and age after birth, too, but this fact hardly means that differences in either height or age are easily changed or narrowed. This holds true for groups as well as individuals. IQ scores have risen during this century for all groups throughout the developed world (Neisser, 1998), but such rises neither logically imply nor have actually been accompanied by any changes in the IQ gaps between those groups. Possible futures aside, the present reality among all existing cohorts is one of stubborn racial gaps in general mental capability.

Turning to the importance of ability differences, dedicated skeptics commonly but mistakenly cite studies of highly restricted populations and narrow skills to refute the conclusion that general mental ability has broad practical importance. For instance, they tend to interpret low observed correlations between the test scores and performance levels of individuals in elite graduate programs or occupations as proof that differences in ability do not matter much and, presumably, that low intelligence is no real handicap. The statistically sophisticated in turn tend to dismiss such interpretations as illustrating mostly the hazards of ignoring the restriction of range in one's samples (e.g., Darlington, 1998). Such sweeping conclusions about importance from atypical slivers of the population are akin to arguing that height makes no difference in basketball because the differences in height among National Basketball Association players are only weakly correlated with the number of points that they score.

Skeptics similarly cite evidence of high levels of development in certain moderately complex skills among people of widely varying IQ levels as evidence that high levels of general ability are not really important in nonacademic settings. They include, among others (Sternberg & Wagner, 1986), studies of racetrack handicapping among devotees (Ceci & Liker, 1986) and homemakers quickly estimating the relative values of differently priced and sized grocery products (Lave, Murtaugh, & de la Roche, 1984). As impressive and useful as these highly specific skills may sometimes be, the claim for the importance of general mental skills does not rest on showing that individuals of below-average ability can never master any moderately complex task, no matter how much they practice. Rather, the claim is that cognitively more able people can more quickly learn and more competently perform the wide range of tasks that life and many jobs demand of them, those tasks often not allowing them extended practice or providing them explicit instruction. This claim emerges from extensive evidence on the validity of mental tests and transferability of general (but not specific) cognitive skills. Studies of isolated and narrow skills do nothing to refute it.

Although the lines of research mentioned above promise to amplify the understanding of mental skills and mental tests, it is only wishful thinking to believe that they make moot the judgment that racial gaps in cognitive test scores represent an important national problem requiring concerted attention.

Confronting Racial Gaps in Skills

As noted, some scholars argue that it is better to foster reassuring lies than to tell unsettling truths. However, acting as if the well-known racial disparities in intellectual skills have no relevance in college, work, and life is folly for tomorrow and no favor today to individuals who are deficient in those skills. Attempting to hide or nullify racial skills gaps by ignoring them at the point of

organizational entry only guarantees that they will become manifested once again as people begin school or work. Tests are not the problem; banishing them is no solution. Skills gaps are the major remaining barrier to racial equality in education and employment, and therein lies the only enduring solution.

Continuing large skills gaps render racial proportionality a premature goal. Proportionality should not even be the criterion for measuring Black progress at this time, as more Black leaders have begun to argue. The National Urban League, for example, resoundingly rejects any notion that Blacks are inherently inferior intellectually, and it certainly does not believe that discrimination has been eradicated (Howard, 1995; Price, 1995). It does emphasize academic development, however, as absolutely central to Black advancement. Stressing that Blacks must lead the academic and economic development of their communities, it sets interim, criterion-related goals for improvement and then tries to help communities to develop the psychological, social, and financial infrastructures to accomplish such development. Progress is not measured relative to Whites but by the distance traversed by the particular individuals, schools, or communities involved.

Selection standards often can be improved, but their opportunistic manipulation for racial reasons will produce only lasting frustration, not enduring solutions. Enduring solutions rest in parsing the problem of skills gaps. There are disproportionately few Blacks among the highly skilled and disproportionately many with deficits severe enough to render them at best marginally employable. Strategies for increasing the number of highly skilled and reducing the number of poorly skilled may differ. Currently, racial preferences help primarily the most skilled Blacks. Enhancing noncognitive traits may be especially useful for improving employability among the least skilled.

Skill levels surely can be increased among all populations, however. Whether or not it ever proves possible to raise or equalize *g* itself, it is possible to improve all students' knowledge levels and highly specific skills, both cognitive and noncognitive. It should also be possible to reverse social trends, especially in family structure, that may stunt the academic development of increasing proportions of minority youth. External assistance can facilitate development efforts within Black communities. However, organizations achieve nothing of lasting value by eliminating valid tests or otherwise reducing achievement standards. Such elimination hinders development while providing only the illusion of progress.

References

Baldwin, J., Kirsch, I. S., Rock, D., & Yamamoto, K. (1995). *The literacy proficiencies of GED examinees: Results from the GED-NALS comparison study.* Washington, DC: American Council on Education and Educational Testing.

Brody, N. (1992). *Intelligence* (2nd ed.). San Diego, CA: Academic Press.

Carroll, J. B. (1987, February). The national assessments in reading: Are we misreading the findings? *Phi Delta Kappan, 68,* 424–430.

Ceci, S. J., & Liker, J. K. (1986). Academic and non-academic intelligence: An experimental separation. In R. J. Sternberg & R. K. Wagner (Eds.), *Practical intelligence: Origins of competence in the everyday world* (pp. 119–142). New York: Cambridge University Press.

Darlington, R. B. (1998). Range restriction and the Graduate Record Examination. [Comments Section]. (1998). *American Psychologist, 53,* 572–573.

Eitelberg, M. J. (1988). *Manpower for military occupations.* Alexandria, VA: Office of the Assistant Secretary of Defense, Force Management, and Personnel.

Glazer, N. (1994, October 31). The lying game. *The New Republic, 16,* 94–95.

Gordon, R. A. (1988). Thunder from the left [Review of *Storm over biology: Essays on science, sentiment, and public policy*]. *Academic Questions, 1,* 74–92.

Gottfredson, L. S. (1986). Societal consequences of the *g* factor in employment. *Journal of Vocational Behavior, 29,* 379–410.

Gottfredson, L. S. (1994a). Egalitarian fiction and collective fraud. *Society, 31,* 53–59.

Gottfredson, L. S. (1994b). The science and politics of race-norming. *American Psychologist, 49,* 955–963.

Gottfredson, L. S. (1996). Racially gerrymandering the content of police tests to satisfy the U.S. Justice Department: A case study. *Psychology, Public Policy, and Law, 2,* 418–446.

Gottfredson, L. S. (1997a). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24,* 13–23.

Gottfredson, L. S. (1997b). Why *g* matters: The complexity of everyday life. *Intelligence, 24,* 79–132.

Howard, J. P. (1995). The third movement: Developing black children for the 21st century. In P. J. Robinson & B. J. Tidwell (Eds.), *The state of Black America, 1995* (pp. 9–29). New York: National Urban League.

Hunt, E. (1995). *Will we be smart enough? A cognitive analysis of the coming workforce.* New York: Russell Sage Foundation.

Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law, 2,* 447–472.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Kirsch, I. S., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993). *Adult literacy in America: A first look at the results of the National Adult Literacy Survey.* Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Kronholz, J. (1998, February 12). As states end racial preferences, pressure rises to drop SAT to maintain minority enrollment. *Wall Street Journal,* A24.

Laurence, J. H., & Ramsberger, P. F. (1991). *Low-aptitude men in the military: Who profits, who pays?* New York: Praeger.

Lave, J., Murtaugh, M., & de la Roche, O. (1984). The dialectic of arithmetic in grocery shopping. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context* (pp. 67–94). Cambridge, MA: Harvard University Press.

Moffitt, T. E., Caspi, A., Harkness, A. R., & Silva, P. A. (1993). The natural history of change in intellectual performance: Who changes: How much? Is it meaningful? *Journal of Child Psychology and Psychiatry, 34,* 455–506.

Neisser, U. (Ed.). (1998). *The rising curve: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101.

Price, H. (1995). Black America, 1994: An overview. In P. J. Robinson & B. J. Tidwell (Eds.), *The state of Black America, 1995* (pp. 1–5). New York: National Urban League.

Reynolds, C. R., Chastain, R. L., Kaufman, A. S., & McLean, J. E. (1987). Demographic characteristics and IQ among adults: Analysis of the WAIS-R standardization sample

as a function of the stratification variables. *Journal of School Psychology, 25,* 323–342.

Rosen, J. (1998, February 23/March 2). Damage control. *New Yorker, 74,* 58, 60, 62, 64–68.

Schmidt, F. (1996, December 10). New police test will be a disaster [Letter to the editor]. *Wall Street Journal,* p. A23.

Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology, 82,* 719–730.

Smith, T. M., Perie, M., Alsalam, N., Mahoney, R. P., Bae, Y., & Young, B. A. (1995). *The condition of education, 1995.* Washington, DC: Office of Educational Research and Improvement. U.S. Department of Education.

Snyderman, M., & Rothman, S. (1988). *The IQ controversy, the media and public policy.* New Brunswick, NJ: Transaction Press.

Spitz, H. H. (1986). *The raising of intelligence: A selected history of attempts to raise retarded intelligence.* Hillsdale, NJ: Erlbaum.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52,* 613–629.

Sternberg, R. J., & Wagner, R. K. (1986). *Practical intelligence: Nature and origins of competence in the everyday world.* New York: Cambridge University Press.

Sticht, T. G., Armstrong, W. B., Hickey, D. T., & Caylor, J. S. (1987). *Cast-off youth: Policy and training methods from the military experience.* New York: Praeger.

Wigdor, A. K., & Garner, W. R. (Eds.). (1982a). *Ability testing: Uses, consequences, and controversies. Part I: Report of the Committee.*Washington, DC: National Academy Press.

Wigdor, A. K., & Garner, W. R. (Eds.). (1982b). *Ability testing: Uses, consequences, and controversies. Part II: Document section.* Washington, DC: National Academy Press.

Zelnick, B. (1996). *Backfire: A reporter's look at affirmative action.* Washington, DC: Regnery.

Received April 7, 1998
Revision received August 5, 1998
Accepted August 16, 1998 ■