

Reconsidering Fairness: A Matter of Social and Ethical Priorities

LINDA S. GOTTFREDSON

Department of Educational Studies, University of Delaware

Current problems of fairness in employment testing do not stem primarily from inadequacies in measurement but from real racial-ethnic differences in the job-related capabilities that the tests often reveal. Moreover, the social and political dilemmas created by such differences are only aggravated by clinging to the hope that either measurement techniques or preferential treatment will provide a satisfactory solution to adverse impact. This paper reviews evidence, first, that it is primarily the *g* (intelligence) factor among mental tests that accounts for their validity in predicting job performance, and second, that black-white differences in *g* are real, large, and stubborn and thus can be expected to lead to especially high levels of adverse impact in mid- and high-level jobs for the foreseeable future. Examples are then provided of how race norming and other forms of preferential treatment designed to prevent adverse impact are counterproductive in the long run, perhaps especially if practiced covertly. Strategies are reviewed by which organizations, knowingly or not, may cultivate the appearance of having done what is generally impossible—avoiding adverse impact while simultaneously maintaining or improving the efficiency and equity of personnel systems—without actually having accomplished that feat. Finally, it is argued that the newer conceptions of fairness, which emphasize group parity rather than individual merit, promise not to bring racial equality but to permanently consign blacks and other favored groups to second-class citizenship. © 1988 Academic Press, Inc.

Concerns over fairness in employment have been driven in recent decades by one fact—that blacks and certain other minorities are “underrepresented” in attractive jobs. Because tests often have been used in evaluating job applicants and because they usually produce adverse impact, that is, lead to the selection of proportionately fewer blacks and Hispanics, they quickly became the focal point of controversy concerning

Keynote address at the conference, “Future Directions of Employment Testing,” held by the Personnel Testing Council of Metropolitan Washington, Washington, DC, July 13, 1988, and revision of a paper presented at the conference, “Fairness in Employment Testing,” held by the Personnel Testing Council of Southern California, Newport Beach, CA, October 1-2, 1987. Correspondence should be addressed to Linda S. Gottfredson, Department of Educational Studies, University of Delaware, Newark, DE 19716.

fairness in employment practices. Despite several decades of research which has largely exonerated tests of the cultural bias charges levied against them, particularly as they concern blacks (Wigdor & Garner, 1982), some people still seek to keep the fairness debate focused on tests and their much-overstated flaws.

As both Schmidt (1988) and Sharf (1988) have pointed out elsewhere in this volume, testing is the wrong focus of debate regarding fairness in employment. We do not have a testing problem so much as we have a social problem brought on by real differences in the job-related capabilities that the tests measure. As has oft been noted, the vulnerability of tests is due less to their limitations for measuring important differences than it is to their very success in doing so. Such is certainly the case where racial-ethnic differences are involved, because the more valid the tests are as measures of general cognitive ability, the larger the average group differences in tests scores they produce. Keeping the spotlight on tests merely forestalls the real debate—how can this society justly and constructively deal with the racial-ethnic differences in ability that will be with us for some time to come?

In this paper I review evidence that current racial-ethnic disproportions in *g* (general intelligence) inevitably lead to adverse impact in many jobs when valid mental tests are used in hiring and promoting workers. I argue that the dilemmas we face with regard to fairness in employment are only aggravated by clinging to the possibility that measurement is the source of and solution to problems of fairness in employment testing. Moreover, race norming and other forms of preferential treatment for blacks and other minorities that are designed to prevent adverse impact are destructive in the long run, perhaps especially so when practiced covertly. I argue that we should not allow the pretense to persist that employers or personnel professionals can do what is usually impossible, namely, avoid adverse impact while simultaneously maintaining or improving the validity and equity of selection systems. Instead, our nation has to take stock of its options, of its choices. That is what should be debated—our ethical and social priorities and the most effective means of pursuing them. We need to examine closely what we consider the good society to be before we have drifted beyond the point of no return away from the fundamental principles upon which this nation was founded.

EVIDENCE ABOUT *g* AND ITS ROLE IN JOB PERFORMANCE

Any assessment of the social problem produced by group differences in mental test scores and the best ways for tackling that problem must be based on a clear understanding of what test scores represent and of why they predict differences in job performance. Attaining such an understanding was the aim of the 1985 Southern California Personnel Testing Council (PTC) conference on *g* and employment testing. The publication

resulting from that conference (Gottfredson, 1986b) reviewed relevant evidence from several subdisciplines, only a few key points of which I shall repeat here.

Evidence about g in Employment

Cognitive tests predict job performance better overall than do other known predictors, and their predictability is due primarily to their measurement of g. As Thorndike (1986) elegantly showed, it is *g*, or the general mental ability factor, that carries the freight of prediction when mental tests are used to predict performance in school and on the job. Special cognitive aptitudes add little to prediction above and beyond this general factor; rather, it is what mental tests measure in common—referred to as the *g* factor—that accounts for the major share of their success in predicting later performance. The overriding importance of *g* holds even when various tests of it bear little or no superficial resemblance to one another or even to the tasks on which performance is being predicted, that is, even when they have little or no manifest content validity. Likewise, Crouse (Gottfredson & Crouse, 1986) reported that it is the *g* factor among ability and achievement tests that accounts for their ability to predict later educational and economic success. Hunter (1986) summarized evidence that specific mental aptitudes (for example, spatial ability) and less cognitive predictors (such as psychomotor ability) sometimes add significantly to the prediction of job performance over and above *g*, but the former have utility only in a few clusters of mid- and high-level jobs and the latter primarily in low-level jobs. Likewise, Gottfredson (1986a) described how measures of experience, training, and psychomotor ability (dexterity and motor coordination) sometimes rival or outperform mental tests as predictors of job performance, but this occurs primarily in the lowest-level jobs.

When considering the full range of jobs in industrialized economies, but particularly the more complex and critical jobs, mental tests are the most important among known predictors of job performance and their utility can be traced primarily to their measurement of *g*. No claim is being made here that cognitive tests or cognitive abilities, general or specific, are the only predictors of job performance. Biodata, vocational interests, and other less-cognitive measures sometimes add substantially to the prediction of certain aspects of job performance beyond that afforded by mental tests; however, this seems to occur primarily when using nontraditional performance *criteria* (for example, “personal discipline” and “military bearing and physical fitness” rather than “job specific core skills” or “general soldiering skills” (Zeidner, 1987, pp. 87–89).

The notion of “multiple intelligences” (Gardner, 1983) is now frequently presented as an alternative to more traditional definitions of intelligence

like *g*. The great appeal of this reconceptualization is that it seems to promise that everyone can be intelligent in some way; that is, it "democratizes the notion of intelligence" (White, 1988). But whether or not we choose to relabel psychomotor ability, social skills, and other less-cognitive abilities as different "intelligences," as is now frequently being done, it does not necessarily follow that all these abilities can be presumed equally useful. As already noted, variation in these other abilities—which were recognized long before Gardner—has not been found as useful as variation in *g*, overall, for predicting variation in the job performance criteria that are typically of most concern.

g is a very general capacity for abstract thinking, problem solving, and learning complex things. Although it is not essential to know the nature of *g* in order to appreciate its practical importance on the job, some knowledge of its established characteristics is helpful for understanding why it is so pervasively important. Briefly, the *g* factor appears to represent an underlying and broad capacity for reasoning, abstract thinking, problem solving, and the related skills by which testing experts define the concept of general intelligence (Snyderman & Rothman, 1986). The *g* factor also corresponds well with lay conceptions of intelligence, and it is correlated with a variety of nonpsychometric phenomena such as reaction time and EEG patterns as well as with genetic phenomena such as the degree of inbreeding depression shown by subtest scores of large batteries (Jensen, 1986).

Understanding what *g* does *not* represent is also important. It does not represent the mere accumulation of bits of knowledge, as though one were filling a jar with marbles and the more one added, by whatever means and at whatever rate, the smarter a person would be. Rather, *g* relates highly to the speed and ease with which people acquire such bits of knowledge, especially complex knowledge, either on their own or through instruction by others. The familiar concept of trainability captures some of what we mean by intelligence, or *g*. As will become clearer below, *g* represents a fundamental capacity which is broader than "academic" ability and it encompasses the currently popular notion of "practical" intelligence (Peters, 1987), even though the latter is often implied to be distinct from and highly independent of academic ability.

g is more important in some jobs than others because their tasks require more problem solving, complex and continuous learning, and other general intellectual skills characteristic of intelligence. Why does *g* predict job performance, especially in jobs that do not seem particularly academic or mental? Hunter (1986) has pointed out that even the most elemental tasks involve mental processes and that most jobs require some judgment in the application of standard procedures. General mental ability does not predict performance equally well in all jobs, but the pattern of predictive validities is neither random nor sensitive to typical local variations

in a job's task content. In fact, validities fall into a very interesting pattern—the higher the job level, the better mental tests predict job performance (Hunter, 1986). As discussed by both Arvey (1986) and Gottfredson (1986a), this pattern is consistent with job analysis data showing, first, that the major factor distinguishing jobs in their task and ability requirements is a general intellectual complexity factor and, second, that higher-level jobs require more reasoning, planning, analyzing, continual learning, and judgment—all being mental skills recognized as aspects of general intelligence. These data provide evidence that there is a *g* factor among jobs themselves; that is, that different occupations can be arrayed from high to low according to the importance of *g* for performing those jobs well. Importance here is reflected by the size of the impact that the *same* difference in *g* has on performance in different jobs and by the minimum levels of *g* effectively required of workers in those jobs. This *g* factor among jobs is highly correlated with both occupational prestige and level of education and training required and so coincides to an important degree with a general desirability factor or status hierarchy among occupations (Gottfredson, 1984).

Evidence about Group Differences in g

There is no debate that different minority groups score differently on mental tests, on the average. The real question is how seriously to take those differences. The answer to this question depends on how real, large, and stubborn the apparent differences in ability are.

Current black-white mental test score differences reflect a black-white difference in the distribution of g. IQ differences between racial-ethnic groups are the rule, not the exception. For example, Jews and Japanese Americans score higher than Anglo whites, and Mexican Americans generally average half a standard deviation and blacks a full standard deviation below Anglo whites (Berryman, 1983; Dearman & Plisko, 1981; Eysenck, 1984; Hennessy & Merrifield, 1978). I focus here on black-white differences because they are the best documented and most at issue in employment testing. My conclusions about adverse impact for blacks relative to whites are equally applicable in principle to other groups who differ in IQ; the major difference lies primarily in the magnitude of the effects, which can be expected to vary according to the size of group IQ difference.

Mental testing experts reluctantly have reached the consensus that the mental tests typically used in educational and employment settings are not biased against blacks (Gordon, 1987; Jensen, 1980; Schmidt, 1988; Wigdor & Garner, 1982). Moreover, the black-white difference is generally acknowledged to be large in a practical sense (Wigdor & Garner, 1982). This point will be graphically illustrated later in this paper.

Although less well accepted yet, evidence is also converging on the

conclusion that mental test score differences between representative samples of blacks and whites primarily reflect differences in *g* across those two populations. One particularly important piece of evidence that black-white differences in mental test scores stem from disproportions in *g* and not, say, from differences in the opportunity to acquire knowledge, is that the magnitude of black-white differences on various mental tests varies with the *g* loading of the test; that is, with its correlation with the first principal factor (Jensen, 1985). Stated another way, tests that distinguish best between more and less intelligent whites or blacks also show the largest differences between the black and white populations. Moreover, black-white differences in test performance at particular developmental ages, say age 8, mirror the differences between younger and older white children (Jensen, 1981). This holds not only for total test scores but also for particular items.

Current black-white differences in g may not be permanent, but they are stubborn. There is considerable disagreement about the origins and malleability of group differences in general mental ability, some of which was reflected in both the previous and the current PTC conferences. On the optimistic side, we know that the environment does affect intelligence to an important degree (Jensen, 1981); there is no absolute proof that any part of the black-white difference is genetic (Jensen, 1981); even if some part of the black-white difference were genetic, it does not preclude remediation with environmental interventions (Plomin, 1987); IQ test scores have risen during this century in many parts of the world on different kinds of intelligence tests and for both blacks and whites (Flynn, 1984; Lynn & Sampson, 1986); and black-white differences on some tests of achievement have narrowed during the last decade or so (Congressional Budget Office, 1986). On the more pessimistic side, the black-white IQ difference has remained remarkably constant at 18 Stanford-Binet IQ points since World War I despite large changes in the relative opportunities and life circumstances of blacks (Gordon, 1980b, in press, Table 2); black-white differences in test scores remain large even after taking account of the major black-white differences in socioeconomic background (Kirsch & Jungblat, 1986); score gains on IQ tests resulting from early interventions do not generalize to other cognitive tasks and usually fade out in a few years, meaning that we do not yet know how to effectively raise individuals' *g* levels (Jensen, 1981); there is no definitive evidence that the widespread rises in IQ scores represent rises in *g* itself; the narrowing of the gap between "urban disadvantaged" and "urban advantaged" students on achievement tests has occurred primarily because of improved performance at the very bottom of the "disadvantaged" distribution (Carroll, 1987), suggesting that black-white differences across the rest of the distribution remain largely unaffected; and even if the black-white gap continues to close at the same rate as it has been on

most tests where the narrowing of the gap has been observed, it will take decades for the gap to disappear (Schmidt, 1988).

The important question in the present context is not whether the current black-white difference in test scores, which appears to be a difference in *g* itself, can or should be decreased in the future. It probably can, and we must try to do so. Denying the reality of the average black-white IQ difference or overstating its tractability will only delay progress toward its elimination.

Rather, the issue is how we should deal with the ability differences between the blacks and whites coming into the labor force now and for at least the near term. A realistic interpretation of the evidence on the stability of intelligence beyond the teen years (Jensen, 1980) is that employers cannot expect to raise the *g* levels of individual workers. Rather, they have to work with what they get. A realistic interpretation of the unexpected resistance of racial-ethnic IQ differences to social change and interventions which were designed to narrow them is that personnel workers should expect to have to deal with these unwelcome differences for at least several decades, if not generations, to come.

Education, training, and experience do not negate the value of g for job performance, and therefore are not likely to negate the impact of black-white differences in g on the job. It is often presumed that ability differences can be compensated for or negated by appropriate education, training, and experience. The "fadeout with experience" theory, which posits that the impact of ability differences will fade as all workers gain experience, and which has been refuted by Schmidt (1988), is one example of such expectations. Likewise, my reading (Gottfredson, in press) of the relevant evidence is that existing education and training strategies do not negate the value of *g* for job performance either. In general, workers with lower *g* levels perform as well as workers with higher *g* levels only when they have some other compensating advantage or *superiority* over the brighter workers. Eventually, as they too acquire experience, brighter workers out-perform equally well-trained or experienced, but less bright, counterparts.

I am not denying the utility of training and experience for improving performance, but only pointing out that education and training strategies are no panacea for black-white differences in workers' *g* levels. Social policies designed to reduce black-white inequalities in employment are often based on the belief that education and training can permanently reduce occupational inequalities despite enduring differences in cognitive ability. This belief appears to arise from misinterpretations of sociological evidence concerning the determinants of workers' occupational status and from misconceptions about employer hiring behavior that I have explicated in the past (Gottfredson, 1985, 1986a).

In short, current black-white differences in test scores must be taken

seriously now, regardless of whatever success we may have in reducing them in the future. They represent real differences in the capacity to learn and perform well a wide variety of job tasks in a wide range of jobs; they have been surprisingly stubborn and so are likely to be with us for some time to come; and their impact on job success is not effectively short-circuited by education, training, or experience. As I shall argue, to ignore them is counterproductive and, in my view, condescending and unfair to blacks.

General Implications of the Evidence about g

Empirical evidence concerning g helps to explain the generalizability of cognitive tests in personnel selection. The patterns of test validities and task requirements described earlier help to explain why validities for mental tests are so generalizable, thereby strengthening the case for validity generalization (Schmidt, 1988) by providing a theoretical explanation for the empirical results. Not only are predictive validities similar for manifestly different cognitive tests (e.g., verbal aptitude, mathematical reasoning, intelligence) but also these tests are predictive for diverse jobs and settings, because most mental tests are primarily measures of *g* and because the *g* factor represents a capability that is useful for performing most kinds of tasks, but especially ones that are complex.

Moreover, the *g* factor describing test scores, on the one hand, and the *g* factor describing jobs and their complexity, on the other hand, together provide an elegant framework for organizing and interpreting the observed variations in validity coefficients, of both cognitive and noncognitive measures, for predicting performance in jobs of different complexity levels. Stated another way, the data on *g* begin to provide a schema for determining the degree to which particular test validities are generalizable to specific other jobs and situations.

The ability to organize existing knowledge meaningfully allows us to be more confident in applying that knowledge effectively in the future. Popular criticisms of tests and validity generalization (Seymour, 1988) might lead one to conclude, mistakenly, that each test and each job must be approached as unique, and thus that we can apply no previous knowledge to new cases. In fact, validity generalization may provide a more accurate validity estimate than would a new study in each setting due to the errors to be expected from the small samples that are typically available for individual test validation studies (Linn & Dunbar, 1986, p. 224).

The evidence about g helps to explain the magnitude and patterns of adverse impact across different jobs and the practical difficulties of reducing adverse impact. The evidence about the pervasive but differential importance of *g* in different kinds of work, taken together with the evidence about the nature and magnitude of group IQ differences, provides a powerful tool for predicting both the degree of adverse impact likely

to be observed from one occupation to another and the practical costs of applying various preferential treatment policies in hiring and promotion in different job settings. This in turn helps to predict which occupations will be the focus of litigation and what kinds of strategies employers may use for avoiding litigation. Finally, as I shall show below, the evidence sheds light on the credibility of employers' or test developers' claims of having avoided or reduced adverse impact without doing damage to the validity of the selection system in question.

AN ANALYSIS OF ADVERSE IMPACT AND STRATEGIES USED TO REDUCE IT

Implications of g for the Magnitude of Adverse Impact

Paradoxically, perhaps, group differences in g guarantee adverse impact in many or most jobs in the nation, although not necessarily in any particular job, for the foreseeable future. Disproportions in *g* levels between the black and white populations do not guarantee that there will be adverse impact in the selection of workers for any particular job, even if that selection is *g* loaded. For example, some employers may be more successful than others in recruiting applicants from the relatively small pool of blacks well qualified for the jobs in question. What is clear, however, is that racial parity across the entire range of jobs is impossible without exercising substantial preferential treatment for blacks in many jobs.

If one accepts the evidence that the black-white difference in *g* levels is stubborn, it also becomes apparent that temporary preferential treatment will not permanently reduce adverse impact. Preferential treatment has been tolerated by many people in large part because it has been presented to them as only a short-term expedient that will not prove necessary in the future. It is easy to see, then, why advocates of temporary preferential treatment might be especially loath to entertain stubborn differences in *g* as a source of race differences in test scores and adverse impact.

Current differences between the black and white IQ distributions are large enough to produce enormous adverse impact when tests are used in a race-neutral manner to hire workers. A fuller appreciation of the predicament facing personnel workers, and the nation itself, can be obtained by looking at black-white disproportions in IQ scores and arraying them along the same *g* dimension that runs through occupations. Figure 1 shows the percentages of the black and white populations falling within different 10-point IQ intervals. There is considerable overlap between the distributions, and both blacks and whites can be found at all IQ levels, but the disproportions are large at all but the IQ 91-100 range. A much smaller proportion of blacks than whites is found above IQ 100; altogether, only about 11% of blacks versus 50% of whites score above

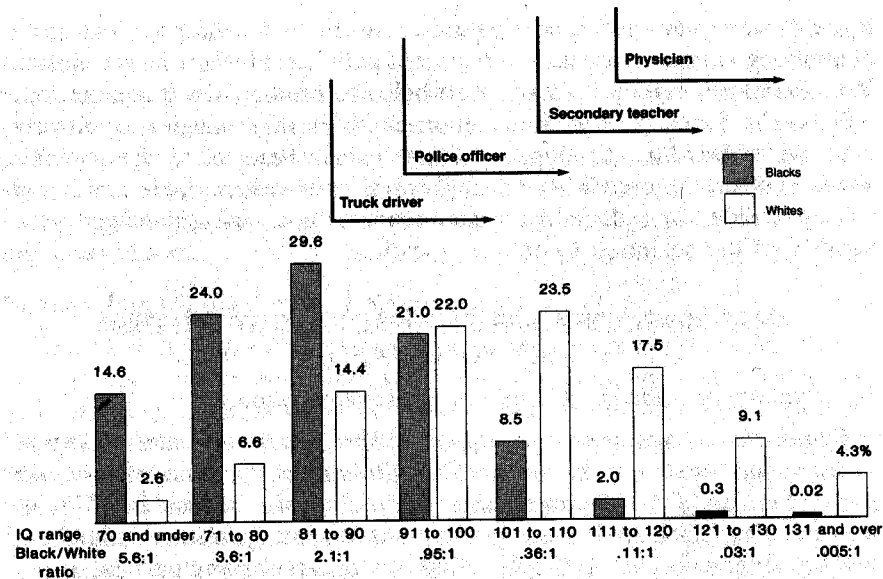


FIG. 1. IQ distributions of blacks and whites as they relate to IQ recruitment ranges for four occupations. Percentages were estimated, using the normal probability table, from means and standard deviations for the black and white populations provided in Gordon (in press, Table 1). Estimated occupational recruitment ranges were obtained from Gottfredson (1986a, Table 2).

IQ 100. The disproportions become especially striking the higher the IQ level considered. Whereas about 4% of the white population falls above IQ 130, only a fraction of 1% of blacks is found above that level, resulting in a minuscule ratio of blacks to whites at that level.

Such dramatic disproportions at the right tail of the IQ distribution may seem implausible unless one is familiar with the properties of IQ distributions, but they are consistent with the large black-white disproportions in other aptitude and achievement test scores. For example, in his book *Choosing Elites*, Klitgaard (1985) reports that the ratio of blacks to whites who scored over 650 on the Graduate Record Quantitative Exam was 1 to 192 in 1978; the ratio of blacks to whites scoring above 700 was 1 to 291. These ratios of, respectively, about 0.5 and 0.3 to 100 are obviously far from parity, which would be represented by a ratio of 12 to 100. The fact that a somewhat smaller proportion of blacks than whites takes such tests obviously cannot begin to account for these minuscule ratios.

If we assume that 650 is a rough threshold for success in a first-rate graduate program in the physical or social sciences (Klitgaard, 1985), the foregoing racial disproportions mean that graduate programs are competing for a very small pool of talented blacks. The ratios just cited

represent 27,470 whites scoring above 650 in that year versus only 143 blacks. The number of blacks scoring above 700 was a mere 50. Harvard alone might have been able to exhaust the entire pool of such top-notch blacks. Blacks obtain more education than do whites of comparable IQ levels (Thomas, Alexander, & Eckland, 1979), so we cannot assume that there exists any untapped reservoir of talented blacks.

Turning to the left tail of the IQ distribution, blacks are much more heavily represented at the low IQ levels than are whites. The ratio of blacks to whites with IQs of 70 or below is almost 6 to 1. The surplus of blacks at the lower IQ levels is consistent with the greater academic difficulties that blacks experience in school, on the average, and with their overrepresentation in classes for the educable mentally retarded (Dearman & Plisko, 1981; Gordon, 1975/1980a, 1980c).

Four IQ thresholds in education will give more concrete meaning to these disproportions (Jensen, 1981). They are probabilistic thresholds only. An IQ of 50 is generally necessary for attending a regular school; an IQ of 75 is generally required for mastering the traditional elementary school curriculum; IQ 105 is generally required for getting grades in an academic curriculum that are good enough for college admission; and IQ 115 is generally required for graduating from a four-year college with grades that would qualify one for admission to a professional or graduate school.

Of more meaning to personnel professionals are the IQ levels that are typical of workers in different jobs. A considerable range of IQ levels is represented in any particular occupation, but the ranges of typical IQ levels in jobs differ systematically by job level (U.S. Department of Labor, 1970, Table 9-2; Stewart, 1947). The higher the job level, the higher the apparent minimum IQ levels. High-level executives, physicians, and other professionals average about IQ 125 and few are found below IQ 115 (Matarazzo, 1972). This would seem to limit the pool of eligibles for this kind of work to about one-quarter of whites but only 1% of blacks. To take several other examples, an estimated minimum for secondary teachers and real estate salespeople is 108; that for police officers and firefighters is 91; and that for truck drivers and meat cutters is 86 (Gottfredson, 1986a). Some occupations draw workers from a somewhat lower IQ range, but they tend not to go below IQ 70-75 (U.S. Department of Labor, 1970, Table 9-2 data on *G* scale scores transformed to the Stanford-Binet metric). Below IQ 70, one is clearly dipping into a pool of hard-to-employ people, for this is the threshold for borderline mental retardation. About 15% of the black population is found at or below this IQ level, which is greater than the proportion of blacks who have IQs above the white IQ average of about 100.

Another way to get a sense of the practical importance of the black-white disproportions in IQ levels is to look at how many jobs recruit

workers from IQ ranges that are centered around the black average of about IQ 85 versus the white average of about IQ 100. This provides a rough estimate of the number of job titles that are available to the bulk of blacks versus the bulk of whites. Of the 441 jobs listed in the General Aptitude Test Battery (GATB) manual according to average GATB G (general intelligence) scale scores (U.S. Department of Labor, 1970, Table 9-2), approximately the middle half of the job titles is readily available to the middle bulk of whites but only the bottom quarter of job titles is available to the middle bulk of blacks.

Once again, no claim is being made here that *g* is the only qualification for work. Some minimum level of IQ is viewed as generally necessary but not sufficient for good performance in an occupation. Indeed, as Schmidt (1988) and others have noted, job performance sometimes can be predicted better while simultaneously *reducing* (but not eliminating) adverse impact when less-cognitive measures are used to supplement cognitive tests in a selection battery. Instead, my aim has been to show how the size of the pool of potential eligibles differs for blacks and whites because of IQ disproportions between the two races. The use of valid less-cognitive supplements can mitigate the adverse impact created by IQ disproportions, but no supplement will eliminate adverse impact in a race-neutral selection process, *if* there are group IQ differences, unless *g* is given no weight in selection or unless the lower scoring group is sufficiently *superior* on some non-*g* component in the selection battery to compensate for its disadvantage with regard to *g*.

Implications of g for Patterns of Adverse Impact

The current black-white disproportions in IQ produce more adverse impact in higher level jobs and make it especially costly to seek parity in those jobs. A few generalizations regarding *patterns* of adverse impact can be drawn from the arguments and data presented above. One is that the higher the job level considered, usually the smaller the pool of potentially eligible blacks is relative to that for whites. Adverse impact can thus be expected to be especially pronounced in hiring for higher level jobs if hiring standards are the same for blacks and whites; similarly the degree of preferential treatment required to avoid that adverse impact will be successively greater at higher job levels.

Also, because of the large size of the average black-white IQ difference, which is over a full standard deviation in representative samples and often close to that among applicants for particular jobs (Jensen, 1977), quite different standards often are required for blacks and whites in order to reduce adverse impact substantially. This also means that many better qualified whites will have to be passed over to get the required ratio of blacks. This can be illustrated with Fig. 1. Most professional and other high-level jobs are filled by people from IQ levels above 110. About 30%

of whites can be found at this or higher IQ levels. In contrast, the IQ level necessary to yield the top 30% of blacks, which would represent racial parity, is IQ 91. This is far below the figure for whites, which was IQ 111. Few professional schools and organizations may actually recruit blacks from such relatively low IQ levels in order to reach full parity, because to do so is to risk stepping across what Gordon (1988) has called the *point of organizational embarrassment*, that is, the point at which an individual's criterion performance is so grossly deficient that it threatens to embarrass or harm the organization.

Another generalization is that especially high degrees of preferential treatment produce especially conspicuous black-white differences in job performance. To the extent that blacks are hired preferentially, as distinct from neutrally, they will typically be clustered among the poorer performers in an occupation, particularly if the preferential hiring had the effect of *raising* the cutoff score for the whites hired at the same time it lowered the cutoff for blacks. (This could happen if both whites and blacks were hired from the top down from separate lists and hiring more blacks meant hiring fewer whites.)

Avoiding adverse impact through preferential treatment in selection for some jobs creates more adverse impact in related jobs. The more preferential treatment there is in higher level jobs, the greater the need will be for employers to exercise preferential treatment in lower level jobs in order to avoid adverse impact there. This is analogous to the ripple effect in institutions of higher education (Klitgaard, 1985). When the more selective universities lower their standards in order to obtain blacks who would otherwise attend somewhat less difficult and less prestigious colleges, those second-tier colleges are then forced to lower their standards in order to recruit blacks who in the past would have gone to third-tier schools. Data provided in Klitgaard's (1985) book illustrates that the degree of preferential treatment accorded blacks in college admissions is often considerable.

In addition, the more that preferential treatment is accorded to blacks at the hiring stage, the greater the preferential treatment that will be required to avoid adverse impact in promotions. Employers who think they are avoiding litigation by exercising preferential treatment in hiring are likely to have a rude shock when promotion time rolls around. Not only are blacks likely to have performed less well on the average if they were selected under lower standards, but jobs higher in a job ladder typically are more *g* loaded, which aggravates the dilemma.

The current surplus of blacks in the "hard-to-employ" segment of the IQ distribution suggests that even substantial preferential treatment would not solve the problem of adverse impact nationwide. Finally, even substantial preferential treatment is unlikely to help many black adults at the bottom of the socioeconomic ladder because many of them will also

be at the bottom of the IQ distribution. As some black conservatives (Loury, 1985) have pointed out, the presence of a large black underclass provides political capital for some politicians and other political players, both black and white, but the underclass does not seem to be the primary beneficiary of the preferential employment policies promoted ostensibly on its behalf. This may be the case because employers are surely reluctant to hire and retain very low ability workers. At some point it becomes more economical for them to mechanize a simple job rather than to employ an unreliable or exceedingly slow worker. It is possible that many of the black underclass adults are, in essence, people who are not even within reach of the bottom rungs of the employment ladder in an industrial society. In view of the mixed success of employment and training programs for the hard-to-employ, employment strategies, whether preferential or not, may not be a significant part of the solution to the problems of the black underclass even though those strategies are now often promoted in its name. Nevertheless, the unemployable segment of the underclass will sometimes be counted as part of the available workforce in calculating degree of adverse impact.

CONSEQUENCES OF THE ACCEPTED IDEOLOGY

Admiring the Emperor's New Clothes

Obviously, the preceding sort of discussion flatly contradicts the accepted ideology which now governs policy-making with regard to fairness in employment (Sharf, 1988). The major tenet of the accepted ideology is that there are no meaningful and enduring racial-ethnic differences in job-related or other capabilities. A corollary is that differences in test scores, education, and employment are due to discrimination against blacks throughout American society. The nation, and whites in particular, are routinely condemned for not having the will or moral fiber to do what is presumably obviously right with regard to blacks. When ability differences are acknowledged to exist and to be significant in practical affairs, they are assumed to be readily remediable. That they have not been remediated so far frequently is attributed to neglect, miserliness, or malevolence on the part of whites.

This ideology grows ever more discrepant with the scientific evidence. Evidence continues to accumulate that the black-white difference in mental test scores is real, stubborn, and of great practical importance, and a consensus that this is so has been developing among testing experts, as indicated by a recent survey (Snyderman & Rothman, 1986). And yet the public increasingly attributes black-white differences in life circumstances to past or current discrimination and decreasingly accords any role to ability differences (Kluegel, 1985). People who dispute the accepted ideology are dismissed by many nonexperts as ignorant or worse. This

new ideology suppresses discussion of racial differences in *g* and their practical consequences, and so allows the miseducation of the public to continue unchecked. It thereby also allows misbegotten public policy to flourish while its proponents proclaim themselves knights battling for a higher moral order rather than being mere political combatants promoting special interests.

But what are the practical consequences of this disjuncture between reality and ideology for employers and personnel professionals? Quite simply, it puts them between a rock and a hard place. They are expected to do the impossible. Except for those lucky employers who manage to obtain more than the usual share of qualified blacks, employers—or their personnel professionals—have to find some workable compromise between increasing minority group representation and the traditional principle guiding personnel work, which is to develop efficient personnel systems that are also viewed as fair and legitimate by the workers involved. The crux of their difficulty is that increasing the representation of blacks generally requires preferential treatment for them, which violates traditional personnel principles regarding merit and traditional American conceptions of fairness toward individuals. Hence, their efforts at compromise typically go underground. The race-norming procedure adopted by the U.S. Employment Service, which is designed to promote parity in job referrals despite large average differences in employment test scores between blacks, Hispanics, and others, was an above-ground compromise. Its very visibility is no doubt why the Employment Service has been challenged by the U.S. Justice Department (Wigdor & Hartigan, 1988).

Proof is hard to come by, but I do not think I am being overly cynical to suggest that many organizations are practicing preferential treatment while cultivating the *appearance* of adhering to professional race-neutral principles. Organizations that cultivate the appearance, knowingly or not, of having accomplished what is in fact usually impossible undermine the efforts of others who would confront the dilemmas more honestly, as well as visit upon themselves and others the destructive consequences of preferential treatment. I will therefore try to raise doubts about all such claims by reviewing strategies by which an organization can project the outward appearance of having done the impossible.

The High Cost of Underground Preferential Treatment

One problem with preferential treatment's going underground is that organizations generally feel compelled to compromise other standards successively in order to keep the original preferential treatment from becoming public, with the end result being a pervasive degradation of standards. Thus, not only are standards lowered for blacks, they are also lowered for nearly everyone. Anecdotes abound, but occasionally there

are documented cases of this. One such case concerns admissions to medical school at Harvard.

During the 1970s Harvard Medical School began reserving 20% of its medical school openings for minorities. The result is a sad and disturbing tale. Keep in mind that Harvard has the special advantage of being able to recruit the best black candidates from the entire country, and therefore other institutions cannot be expected to fare any better than did Harvard. Gordon's (1988) review of Bernard Davis's (1986) book *Storm Over Biology: Essays on Science, Sentiment, and Public Policy* recounts the silent erosion of standards that Davis witnessed:

The new admissions standard for minority applicants had a domino effect on various policies that had evolved over the years to enable the school to monitor its product and maintain its commitment to excellence. Because black students experienced their greatest difficulty in basic science courses, it was suggested that the "long tradition of building on these courses as a foundation for clinical training might have been wrong: perhaps one really did not need to be competent in science in order to be a good physician." Letter grading was replaced by the less informative pass-fail criterion, and incompletes were rendered invisible on student records once the missing coursework had been made-up. Such changes made it easier for the dean to claim that performance records of minority students were indistinguishable from those of other graduates. Departments were pressed to permit repeated re-examinations for failing students, "and inevitably these examinations became less demanding." As a by-product, the standards for passing crept downward for all students.

Before long, the dean's office discontinued yearly reporting of the school's students in the National Board Examinations, until then a ritual. Eventually, the faculty came to rely on passing the National Board Examinations as evidence that its standards had not declined too far, although Harvard would have considered such a criterion excessively permissive for its students in the past. But the National Board Examinations are renormed each year, Davis informs us in another essay, "and so the absolute norm for passing is necessarily lowered by any nationwide increase in admission of students with substandard academic qualifications." . . . (At Harvard) a failing student could retake the National Board Examinations five times, but eventually that anemic standard was itself waived and a diploma awarded in the case that at last caused Davis to publish a 1976 guest-editorial in the *New England Journal of Medicine* in which he sounded the alarm. The point of organizational embarrassment had finally been passed. . . .

At Harvard, the final domino was the tradition of *veritas*. In jeopardy all along as the faculty was systematically deprived of objective feedback on the performance of students, the tradition collapsed catastrophically as the administration maneuvered to contain the embarrassment caused by Davis's principled whistle-blowing, as the dean sent out a letter to all medical schools denying that standards had been lowered at Harvard and issued a misleading press release castigating Davis, as Davis's colleagues abandoned him publicly, as blacks debated whether or not he was a racist, and as the *Harvard Crimson* and Richard Lewontin rushed to depict him as indeed a racist who questioned the ability of all black medical students if not all blacks (pp. 85-87).

As this case shows, it may be only at the point of organizational

embarrassment, when standards have already been intolerably degraded, perhaps irreparably, that the process is belatedly exposed and we learn what has happened. I suspect that a similar process of degradation in personnel selection and promotion standards has set in throughout the country, especially in metropolitan areas. It may be proceeding faster in public institutions than private ones, but is probably endemic throughout all sectors. One can only speculate at this point how far systems are likely to be degraded before attempts are made to halt or reverse the process, but we must not be sanguine about prospects for easy recovery.

A permanent change in the quality of workers flowing into an occupation changes the very nature of the occupation over time (Gottfredson, 1985). Hiring blacks under lower standards will result in productivity losses to the extent that less qualified blacks displace more qualified applicants. Although the immediate impact on productivity may be important and the longer-term shift in the nature and organization of the occupation's tasks perceptible, they are nothing compared to the dramatic systemic effects when organizations try to hide an otherwise noticeable average black-white difference in performance by lowering standards for *all* their workers or students.

This effect is analogous to the consequences of different selection strategies for reducing adverse impact that Hunter and his colleagues (Hunter, 1986; Hunter, Schmidt, & Rauschenberger, 1984) have described. Using explicit racial quotas but still hiring the highest scorers of both races often exacts a much smaller price in immediate productivity losses than does setting a low cutoff for both races and then hiring from the remaining applicant pool according to other criteria. I cite this merely to illustrate that underground preferential treatment is more expensive than usually realized. I do *not* mean to imply that explicit quotas are acceptable. As I shall argue later, preferential treatment by race is not a medication that will cure our ills, whether it be administered openly or not, but rather it is a cancer on the body politic.

Strategies for Underground Preferential Treatment Which May Now Be in Use

Arguments about fairness in testing focus primarily on the hiring process, so attention will be limited here to hiring standards. A variety of race-conscious and race-neutral selection strategies have been proposed by researchers, the courts, and others. All of the race-conscious strategies are known to result in lower worker productivity than do the most efficient race-neutral strategies (Wigdor & Hartigan, 1988). Here I describe a few more surreptitious strategies for reducing adverse impact which upgrade selection but without the *obvious* appearance of doing so. They are ways of seeming to do the impossible.

If black applicants have lower average scores than white applicants

on *g*-loaded tests, there are at least four options for reducing adverse impact without adopting double standards: (a) reduce the *g* loading (the difficulty) of the test itself; (b) replace the test with a less *g*-loaded alternative; (c) reduce a *g*-loaded test's weight in the overall selection process; or, perhaps most insidiously of all, (d) reduce the *g* loading of the criterion against which the test is validated. If instituted without a good rationale—for example, a conscious decision that the new performance criterion really is more representative of the organization's performance goals or that the new selection battery is at least as valid as the old—then these options result in less efficient selection for all workers.

Making the test easier (less g loaded). With regard to the first option, the *g* loading of an employment test can be reduced by eliminating the more *g*-loaded items. These will often be the items on which blacks and whites show the largest differences. For example, one can often obtain easier test items generated from a job analysis by picking job tasks that are frequent rather than critical. The items will still be manifestly job related and the test will thus appear to be content valid and professionally defensible. If data on predictive validity are not made available, any degradation in predictive validity will not be readily apparent. Now that the Supreme Court's decision in the *Watson* case may have eased the burden of proof on employers to show predictive validity, which I think is reasonable, it may be easier to hide any degradation in test utility. Careful attention to content validity can improve the predictive validity of a test, which is one reason that it is important in test development, but superficial appeals to manifest content validity are no guarantee that the test has much predictive validity.

Switching to an easier (less g-loaded) "alternative." With regard to the second option, switching from a mental test to educational credentials, biodata, interviews, subjective measures, and other less *g*-loaded alternatives would generally reduce adverse impact. If the switch is not accompanied by data showing that the alternative is at least as valid as the abandoned test, then one might reasonably suspect that selection has been degraded. Even when the alternative is just as valid as a cognitive test, performance usually will be better predicted when the new predictor is used as a *supplement* to rather than as a replacement for the cognitive test.

Giving less weight to the g-loaded predictors in a selection battery. Using test scores differently can also decrease the weight of *g* in the selection process, which is the third option mentioned above. For example, rather than selecting the applicants with the highest test scores, one could use tests only to screen out the lowest scoring applicants and then make the final selections according to other criteria. As Schmidt and his co-workers (Hunter et al., 1984; Schmidt, 1988) have pointed out, much of a test's efficiency is lost by setting a low cutoff rather than ranking

scores and hiring from the top down. Supplementing cognitive tests with other predictors in an efficiently applied selection battery sometimes reduces adverse impact while also increasing predictive validity, as was discussed earlier, but this happy circumstance must be demonstrated, not assumed, because it occurs much less often than expected. Without such evidence, one cannot be sure that adding the supplement did not lower the utility of the selection battery.

Changing the performance criterion to one that, compared to the previous criterion, is less well predicted by mental tests and better predicted by less-cognitive predictors. Finally, one can substitute performance criteria so as to make less-cognitive predictors look better. One long-standing hope among many people had been that switching from performance ratings to objective hands-on performance tests as criteria would reduce the predictive validity of mental tests. To date, however, evidence suggests that the predictive validity of mental tests *increases* when more objective performance criteria are used, at least for middle- and high-level jobs (Hunter, 1986; Schmidt, 1988; Zeidner, 1987). This suggests that subjective criteria are generally less *g* loaded and that their adoption in favor of objective criteria, appropriately or not, might reduce the predictive validity of mental tests relative to alternative predictors.

Criteria for academic performance provide a clear illustration of how performance criteria can be altered to diminish the apparent importance of *g* as a predictor. Specifically, the validity of mental tests for predicting academic performance can be reduced by substituting more subjective criteria such as grades for standardized achievement tests as the criterion measure. The validity of mental tests for predicting grades themselves can be further reduced by including nonacademic subjects in calculating the criterion scores. Gordon (1980c) illustrated how some critics of tests have used this strategy to impugn the fairness of standardized tests of ability for minority children in elementary school. Comparable possibilities for personnel selection can be imagined by reviewing the matrix of validities for various combinations of predictors and criteria provided by Zeidner (1987, Tables 31–34).

There are no formal procedures for justifying criterion measures comparable to those for validating predictors, so it is harder both to challenge and to defend criteria than it is to evaluate predictors. All shifts in performance criteria should be accompanied by a persuasive defense that extends beyond consideration of reductions in adverse impact.

Selections from the Lucky Few

A reduction of adverse impact does not necessarily mean that selection standards have been lowered for any racial group, because black applicant pools may have been large enough or comparable enough across races for employers to select the desired ratio of blacks to whites without

instituting double standards or degrading standards for both races. It is instructive, however, to review the circumstances in which this fortunate outcome might occur.

A major possibility is that the organization simply has outbid other organizations in the small market of well-qualified blacks. For example, by virtue of its prestige, Harvard is in a favorable position relative to other universities when recruiting students and faculty. Less prestigious universities may be able to outbid other institutions by offering especially attractive salaries or reduced work loads. Second, the organization's geographic location may be favorable for obtaining good applicants. Where the relevant workforce is national, this may mean being located in an urban area which provides an attractive social life for blacks, versus being located in a less urban and hence whiter setting. On the other hand, if the labor force deemed relevant for affirmative action purposes happens to be the local one, then being located in less urban areas may prove to be an advantage by providing a low standard of parity that is easily met by recruiting blacks from further away than whites.

Other circumstances affect applicant pools, but these examples are sufficient to make the point that one employer's ability to escape the tradeoff between group representation and degraded or double standards is no sign that all employers can do so. In fact, it usually means that fewer well qualified blacks are available to the next employer, which only worsens that employer's dilemma. Such success does not reflect any magic formula for overcoming racial differences in test scores. To return to the Harvard example, Harvard may be able to recruit the greater part of the pool of black students scoring high on the SAT or GRE and present itself as a model of affirmative action, but other colleges and universities may find those claims to moral superiority a bit hollow and self-serving as they go about trying to recruit viable black candidates for their own programs. And recall from the medical school example that even Harvard does not seem to be able to escape rather dramatic tradeoffs in its affirmative action programs.

The Ethics of Rejecting the New Ideology

Some people argue that it harms blacks to talk about a possible average racial difference in intelligence, whether it is partly genetic or not, and that it is better to promote what they view as a useful fiction. But as explained elsewhere (Gottfredson, 1987) it endangers blacks more, and the nation too, if we fail to confront and come to grips with the situation. Promoting the fiction is not harmless, as many people seem to assume. Just because academics, politicians, and others may refuse to discuss race differences in *g* publicly is no guarantee that the lay public will not notice, and perhaps even dangerously misconstrue, the inevitable dif-

ferences in real-world performances that follow from racial-ethnic disproportions in *g*.

SHIFTING CONCEPTIONS OF FAIRNESS AND THEIR IMPLICATIONS

Historically, the major task before personnel researchers and professionals has been to find ways of improving worker productivity. In view of recent concerns about international competitiveness and in view of the demographic trends now transforming this nation's labor force (Johnstone & Packer, 1987), one might suppose that personnel workers would be confident about their potential contributions to the welfare of the nation and of receiving support for their efforts. Not so. Like educators, who also sit at major switch points in society where individuals are evaluated and selected into desired social positions, they are often expected to transform unequal inputs into more equal outputs, all the while promoting excellence. That the task may be impossible does not spare them criticism for failure or insistence on success.

The tradeoff between equality and efficiency or excellence has long been recognized among economists (Okun, 1975) and other social scientists (Gardner, 1984). The balance that is struck, and the means by which it is sought, reflect the most fundamental principles of a nation. A look at the current debate about the equality-performance trade off in personnel selection helps to illustrate that our polity is undergoing a momentous but underappreciated transformation in its conceptions of fairness and its vision of the good society. The bargain now being struck in personnel selection with regard to fairness both reflects and reinforces the larger political transformation, so it is also instructive to look at the role of personnel scientists and professionals in fashioning that bargain.

Personnel selection research and practice once proceeded from two basic principles: first, it is good to promote higher levels of worker performance and productivity, and, second, it is good to recognize and reward merit, which primarily meant skill, effort, and the taking of responsibility. Largely because of the American emphasis on individualism, but also because of practical difficulties in studying the performance of groups rather than individuals, the emphasis was on promoting and rewarding the performance of *individuals*. Fairness was judged according to whether individuals were in fact treated according to their merits. The controversy about possible racial bias in employment testing epitomizes this conception of fairness—are people of all races judged equally, in a color-blind manner, according to the relevant skills, abilities, and experience they have to offer employers? Justice meant both fairness to individuals in the distribution of opportunities and rewards and also an appropriate degree of difference in rewards, that is, that differences in reward be large enough to provide incentives but not so large as to create resentment and hardship among the less successful workers. To some extent, it has

always been viewed as prudent as well as just for employers to temper the principle of reward for merit with a consideration of workers' needs. Increasing pay levels with seniority is one such example. Although senior workers are not necessarily better workers, turnover is quite costly to the employer.

Concerns about persisting racial inequalities have changed all this. Once it became clear that good-faith applications of these principles did not eliminate racial differences in employment, indeed, that they hardly dampened them at all, the principles themselves came under attack. Where once these principles were taken for granted, they now must be defended. Where we once spoke simply of "rights," which were presumed to inhere in individuals, many people now distinguish between individual and group rights and accord the latter higher standing. Where once it was assumed that hiring according to individual merit was most fair, it is now claimed that some minority groups have a right to a larger share of certain jobs—not to recognize previously unappreciated merit, but to compensate for their greater disadvantages, handicaps, and needs. To promote productivity and excellence was once believed to promote the common good, but people have begun debating how much we should sacrifice such laudable goals in order to promote so-called group rights in the name of a dubiously redefined common good.

Several decades ago, personnel selection researchers worked to develop selection systems that were *unbiased* and *more valid* predictors of job performance; this was perceived as promoting greater fairness for everyone. It has since become obvious that improving selection along these lines usually *increases* adverse impact, which therefore puts such activity completely at odds with the new conceptions of "fairness to groups" that figure so strongly in current debates about personnel testing (Wigdor & Hartigan, 1988). In the name of fairness, ironically, there thus has developed effective pressure to reverse traditional personnel practice, that is, to develop and use tests that are *biased* (to favor blacks) and *less valid*.

The standard for fairness now is often group equality of results, regardless of differences in inputs, which are often asserted to be either nonexistent or inconsequential. Perhaps for lack of a more substantial argument to defend equality of results, "diversity" has become an overriding virtue. Individual rights used to permit true diversity; now those rights must be curtailed to achieve a carefully "balanced" workforce that regards all individuals within specified racial-ethnic groups as fungible for purposes of calculating and enforcing diversity. To illustrate how ascendant these new principles have become, consider the attitude of the University of California's president as reflected in the headline of one news article (Smith, 1988, p. A4): "Bias charge frustrates UC chief; cites common good as race is officially made factor in [college] admission."

I believe that this shift in received opinion about what is in the common good and its relatively unchallenged progress, despite considerable private resentment, can be attributed in large part to our exaggerated fear of confronting group differences in intelligence and our consequent effort to cope with them without even acknowledging their existence. A large segment of the nation denies that such differences exist despite ample evidence to the contrary; another large segment finds it expedient to refrain from questioning the new mythology. For lack of public dissent, which has been effectively muffled, the new principle of preferential treatment to secure "group rights" gains legitimacy through mere repetition and thoughtless, or expedient, capitulation.

In my experience, not only is the fear of confronting group IQ differences grossly exaggerated, but also the etiquette of ignoring them is condescending and harmful to its ostensible beneficiaries. As described elsewhere (Blits & Gottfredson, 1988), the terrible irony is that the shift in conceptions of fairness that seems to promise fulfillment of our current quest for racial equality will actually institutionalize its opposite. The continuing focus on test bias serves only as a smokescreen that allows this tragedy to unfold largely unnoticed.

If all blacks are set apart as eligible for special treatment without regard to their individual needs and capabilities, without regard to their individual strengths and weaknesses, then all will wear the yoke of inferiority. This will be especially so, despite the new mythology, if the IQ disproportions remain unremediated and continue to manifest themselves in obvious average differences in performance. Many black individuals will never develop to their potential. Others will never reap the gratification of knowing for sure that they succeeded on their own merits—and that other people know that too—in a society where merit may be envied but has always been respected. But strong or weak, noble or base, the fates of all black individuals will be bound together inextricably as blacks, Hispanics, and other officially designated groups compete for their "appropriate" share of an ever shrinking pie.

Gordon (1981) has described the political choice before us as one between "liberal pluralism" and "corporate pluralism." The former essentially represents the liberal tradition on which the nation was founded. The latter gives formal recognition to categories of people based on race and ethnicity, and it distributes political power and economic rewards according to formulas based on group representation. Group equality of condition is favored over meritocracy and equality of opportunity. Because group membership plays such a large role in access to power and economic rewards, there is distinct pressure for people to marry and associate primarily with others in their own group. Corporate pluralism values and perpetuates cultural differences, it favors institutional bilingualism or multilingualism, and "its emphasis on group identity and group rights

makes it less insistent on the principle of free access in travel or residence" (p. 186).

Blits and Gottfredson (1988) have described that choice more starkly by drawing out its implications. The current drift toward "corporate pluralism," which is balkanizing the country by racial-ethnic group, is taking us back to feudalism. Under feudalism there existed distinctly different social stations, each of which was subject to different sets of laws which were adjudicated in different courts. One's station in life was determined completely by one's status at birth. One's rights, responsibilities, and opportunities were in turn determined by one's station, and one could be the equal only of fellow members of one's group. Relations among the groups were regulated in no small measure by brute force and naked power. There was neither liberty nor equality.

CONCLUSIONS

A knowledge of current racial-ethnic disproportions in intelligence is sobering, but is no cause for despair or inaction. There is widespread agreement that greater equality in life circumstances by race is desirable; people differ largely in the remedies they believe will be effective and tolerable. The search for effective remedies for large racial-ethnic inequalities is more likely to be successful if the problem is accurately diagnosed. This paper, like Schmidt's (1988), is designed to help inform the search for solutions, not to short-circuit it. That I reject preferential treatment by race reflects my judgment that it will do great harm in the long run to the nation and to the favored groups themselves.

Due to our collective reluctance to even acknowledge the existence of stubborn and consequential group differences in intelligence, we have hardly begun to consider our possible options for coping with or ameliorating them. We must not fall prey to an undue pessimism born of ignorance, but should examine more carefully what our goals and options are. In addition, our exploration of alternatives may be more productive if we reconsider common attitudes toward ability, equality, and human dignity that may be hindering rather than furthering our efforts to promote racial justice. In this spirit, I suggest that the following five principles guide deliberations about reducing racial-ethnic inequalities in employment.

1. *Remember that we are all in the same boat together.* If for no other reason than this, we should all be concerned about fashioning selection systems—and a society—that people from all racial-ethnic groups, majority and minority, consider fair and legitimate.

2. *Multiple strategies for reducing employment differences should be pursued, only some of which will involve employment policy per se.* Neither employers nor their tests should be expected, by themselves, to eliminate racial differences in employment. Employers must be fair, but as I have outlined in this paper, unrealistic expectations are coun-

terproductive. One contribution employers can make is to search for valid supplements to mental tests. Not only do valid supplements serve employers' interests by providing more efficient selection systems, but they often reduce adverse impact. Other points of intervention include reducing the IQ differences between races and improving training for lower IQ individuals, so research and development in these areas might be pursued more vigorously.

3. *Parity in employment is not necessary for reducing differences in life circumstances.* Employment policy, like family and welfare policy, is but one among multiple strategies for improving people's life circumstances.

4. *Judgments about the dignity or worth of any group or member thereof, or the justness of a society, should not stand or fall according to degree of racial parity in employment or other individual social outcomes.* Demanding such parity is unrealistic; demanding it in the name of securing dignity is thus tantamount to denying dignity to the intended beneficiaries when such demands cannot be fulfilled.

5. *Just as all groups and individuals can be full partners in promoting productivity and the common welfare, whatever their ability levels, so too can—and should—people from all groups be full partners in finding remedies for the causes and consequences of group ability differences.* Honest, nonpatronizing, nondefensive, and nonaccusatory discussion of problems and possible solutions is essential for such a communal effort.

REFERENCES

- Arvey, R. (1986). General ability in employment: A discussion. *Journal of Vocational Behavior*, 29, 415-420.
- Berryman, S. E. (1983). *Who will do science? Minority and female attainment of science and mathematics degrees: Trends and causes*. New York: The Rockefeller Foundation.
- Blits, J. H., & Gottfredson, L. S. (1988). *Race-norming: Equality at last, or lasting inequality?* Unpublished manuscript.
- Carroll, J. B. (1987). The National Assessments in Reading: Are we misreading the findings? *Phi Delta Kappan*, 68, 424-430.
- Congressional Budget Office (1986). *Trends in educational achievement*. Washington, DC: The Congress of the United States, Congressional Budget Office.
- Davis, B. (1986). *Storm over biology: Essays on science, sentiment, and public policy*. Buffalo, NY: Prometheus Press.
- Dearman, N. B., & Plisko, V. W. (1981). *The condition of education*. Washington, DC: National Center for Education Statistics.
- Eysenck, H. J. (1984). The effect of race on human abilities and mental test scores. In C. R. Reynolds & R. T. Brown (Eds.), *Perspectives on bias in mental testing* (pp. 249-292). New York: Plenum.
- Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, 95, 29-51.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- Gardner, J. (1984). *Excellence: Can we be equal and excellent too?* New York: Harper.

- Gordon, M. M. (1981). Models of pluralism: The new American dilemma. *The Annals of the American Academy of Political and Social Science*, 454, 178-188.
- Gordon, R. A. (1975/1980a). Examining labelling theory. In W. R. Gove (Ed.), *The labelling of deviance: Evaluating a perspective* (2nd ed., pp. 111-174). Beverly Hills, CA: Sage.
- Gordon, R. A. (1980b). Implications of valid (and stubborn) IQ differences: An unstatesmanlike view. *Behavioral and Brain Sciences*, 3, 343-344.
- Gordon, R. A. (1980c). Labelling theory, mental retardation, and public policy: Larry P. and other developments since 1974. In W. R. Gove (Ed.), *The labeling of deviance: Evaluating a perspective* (2nd ed., pp. 175-225). Beverly Hills, CA: Sage.
- Gordon, R. A. (1987). Jensen's contributions concerning test bias: A contextual view. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (Falmer International Master-Minds Challenged). Sussex, England: Falmer Press.
- 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.
- Gordon, R. A. (in press). IQ-commensurability of black-white differences in crime and delinquency. *Personality and Individual Differences*.
- Gottfredson, L. S. (1984). *The role of intelligence and education in the division of labor* (Report No. 355). Baltimore, MD: The Johns Hopkins University, Center for Social Organization of Schools.
- Gottfredson, L. S. (1985). Education as a valid but fallible signal of worker quality: Reorienting an old debate about the functional basis of the occupational hierarchy. In A. C. Kerckhoff (Ed.), *Research in sociology of education and socialization* (Vol. 5, pp. 123-169). Greenwich, CT, JAI Press.
- Gottfredson, L. S. (1986a). Societal consequences of the g factor in employment. *Journal of Vocational Behavior*, 29, 379-410.
- Gottfredson, L. S. (Ed.) (1986b). The g factor in employment [Special issue]. *Journal of Vocational Behavior*, 29(3).
- Gottfredson, L. S. (1987, August). *Breaching taboos: A personal perspective*. Paper presented at the annual meeting of the American Psychological Association, New York.
- Gottfredson, L. S. (in press). Intelligence versus training: Job performance and black-white occupational inequality. *Personality and Individual Differences*.
- Gottfredson, L. S., & Crouse, J. (1986). Validity versus utility of mental tests: Example of the SAT. *Journal of Vocational Behavior*, 29, 363-378.
- Hennessy, J. J., & Merrifield, P. R. (1978). Ethnicity and sex distinctions in patterns of aptitude factor scores in a sample of urban high school seniors. *American Educational Research Journal*, 15, 385-389.
- Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, 29, 340-362.
- Hunter, J. E., Schmidt, F. L., & Rauschenberger, J. (1984). Methodological, statistical, and ethical issues in the study of bias in psychological tests. In C. R. Reynolds & R. T. Brown (Eds.), *Perspectives on bias in mental testing* (pp. 41-99). New York: Plenum.
- Jensen, A. R. (1977). An examination of culture bias in the Wonderlic Personnel Test. *Intelligence*, 1, 51-64.
- Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- Jensen, A. R. (1981). *Straight talk about mental tests*. New York: Free Press.
- Jensen, A. R. (1985). The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences*, 8, 193-219.
- Jensen, A. R. (1986). g: Artifact or reality? *Journal of Vocational Behavior*, 29, 301-331.
- Johnstone, W. B., & Packer, A. H. (1987). *Workforce 2000: Work and workers for the 21st century*. Indianapolis, IN: Hudson Institute.
- Kirsch, I., & Jungblat, A. (1986). *Literacy: Profiles of America's young adults* (Report No. 16-PL-02). Princeton, NJ: National Assessment of Educational Progress.

- Klitgaard, R. (1985). *Choosing elites*. New York: Basic Books.
- Kluegel, J. R. (1985). "If there isn't a problem, you don't need a solution:" The bases of contemporary affirmative action attitudes. *American Behavioral Scientist*, 28, 761-784.
- Levin, H. M. (1985). *Educational reform for disadvantaged students: An emerging crisis*. Washington, DC: National Education Association.
- Loury, G. C. (1985, Spring). The moral quandary of the black community. *The Public Interest*, No. 79, 9-22.
- Linn, R. L., & Dunbar, S. B. (1986). Validity generalization and predictive bias. In R. A. Berk (Ed.), *Performance assessment: Methods and applications*. Baltimore: Johns Hopkins University Press.
- Loury, G. C. (1985, Spring). The moral quandary of the black community. *The Public Interest*, No. 79, 9-22.
- Lynn, R., & Sampson, S. (1986). The rise of national intelligence: Evidence from Britain, Japan, and the U.S.A. *Personality and Individual Differences*, 7, 23-32.
- Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). Baltimore: Williams & Wilkins.
- Okun, A. M. (1975). *Equality and efficiency: The big tradeoff*. Washington, DC: Brookings Institution.
- Peters, R. J. (1987). *Practical intelligence*. New York: Harper & Row.
- Plomin, R. (1987). Behavioral genetics and intervention. In J. J. Gallagher & C. T. Ramey (Eds.), *The malleability of children* (pp. 15-24). Baltimore: Brooks.
- Schmidt, F. L. (1988). The problem of group differences in ability test scores in employment selection. *Journal of Vocational Behavior*, 33, 272-292.
- Seymour, R. T. (1988). Why plaintiffs' counsel challenge tests, and how they can successfully challenge the theory of "validity generalization." *Journal of Vocational Behavior*, 33, 331-364.
- Sharf, J. C. (1988). Litigating personnel measurement policy. *Journal of Vocational Behavior*, 33, 235-271.
- Smith, J. L. (1988, May 21) Bias charge frustrates UC chief: Cites common good as race is officially made factor in admission. *The Sacramento Bee*, p. A4.
- Snyderman, M., & Rothman, S. (1986, Spring). Science, politics, and the IQ controversy. *The Public Interest*, No. 83, 79-98.
- Stewart, N. (1947). Relationship between military occupational specialty and Army General Classification Test standard score. *Educational and Psychological Measurement*, 7, 677-693.
- Thomas, G. E., Alexander, K. L., & Eckland, B. K. (1979). Access to higher education: The importance of race, sex, social class, and academic credentials. *School Review*, 87, 133-156.
- Thorndike, R. L. (1986). The role of general ability in prediction. *Journal of Vocational Behavior*, 29, 332-339.
- U.S. Department of Labor. (1970). *Manual for the USTES General Aptitude Test Battery*. Washington, DC: U.S. Government Printing Office.
- White, S. W. (1988). Opportunity and intelligence. *The Phi Kappa Phi Journal*, 68, 2-3.
- Wigdor, A. K., & Garner, W. R. (Eds.) (1982). *Ability testing: Uses, consequences and controversies: Pt. 1. Report of the committee*. Washington, DC: National Academy Press.
- Wigdor, A. K., & Hartigan, J. A. (Eds.) (1988). *Interim report: Within-group scoring of the General Aptitude Test Battery*. Washington, DC: National Academy Press.
- Zeidner, J. (1987). *The validity of selection and classification procedures for predicting job performance* (IDA Paper P-1977). Alexandria, VA: Institute for Defense Analysis.

Received: August 10, 1988