Journal of Industrial Psychology, 1996, 22(1), 1-6
Tydskrif vir Bedryfsielkunde, 1996 22(1), 1-6

THE SOCIO-POLITICAL CONTEXT OF THE APPLICATION OF FAIR SELECTION MODELS IN THE USA

G K HUYSAMEN

Department of Psychology
University of the Orange Free State

OPSOMMING

'n Vorige artikel het die psigometrika onderliggend aan verskeie billike keuringsmodelle wat in die laat sestigerjare, vroeë sewentigerjare in die Verenigde State van Amerika voorgestel is, behandel. Die doel met die onderhawige artikel is om 'n oorsig te verskaf van die daaropvolgende geskiedenis van die toepassing van daardie modelle in personeelkeuring in daardie land, en om die implikasies daarvan vir die Suid-Afrikaanse situasie te belig. Omdat die aangeleentheid van billike keuringsmodelle verband hou met die kwessie van regstellende aksie, word 'n bondige geskiedenis van hierdie kwessie soos dit op personeelkeuring van toepassing is, ook verskaf. Sleutel-uitsprake van die Amerikaanse Hooggeregshof wat betrekking het op hierdie aangeleentheid word ook beskou. Die beperkte toepassing van hierdie billike keuringsmodelle kan toegeskryf word aan die heersende sosio-politieke konteks wat die voorkeurbehandeling van bepaalde groepe voorstaan, maar wat huiwerig is om die besonderhede en perke van sodanige behandeling te spesifiseer.

ABSTRACT

In an earlier article, the psychometrics of various fair selection models that had been proposed in the United States of America in the late 1960s, early 1970s were presented. The purpose of the present article is to discuss the subsequent history of the application of these models in personnel selection in that country and to view its implications for the South African situation. Because the question of fair selection models ties in with the issue of affirmative action, a brief history of this issue as it pertains to personnel selection is also given. Key decisions of the American Supreme Court that have a bearing on this matter are also reviewed. The failure to widely apply these fair selection models may be attributed to the prevalent socio-political context which favours the preferential treatment of certain groups but is hesitant to specify the particulars and limits of such treatment.

Like their counterparts in the United States of America, South African employers face a dilemma when it comes to selection and promotion decisions. Because our American colleagues have grappled with this problem for at least the past quarter of a century, it may be instructive to examine how they have coped with it. Their experts generally agree that, for most jobs, the mean on selection tests for whites is about one standard deviation higher than that for black (or African-) Americans (Gottfredson, 1994; Sackett & Wilk, 1994). Moreover, research overwhelmingly suggests that these group differences are not due to bias in the tests as predictors of job performance, but rather reflect real differences in work-related abilities and skills (Schmidt, 1988). As a result, the colour-blind use of these tests to appoint employees would promote workforce productivity. However, the test-score difference between the groups involved is of such a magnitude that selection in terms of these (valid, unbiased) tests would result in the selection of blacks in numbers that are disproportionately smaller than their numbers in the applicant pool. This 'perverse truth', as Gottfredson (1994) calls it, will persist as long as different groups continue to differ (to the extent that they do presently) in their scores in valid, unbiased selection tests. In the meantime, the employer has to walk a tightrope between the objective of higher productivity, on the one hand, and that of righting past social wrongs, on the other hand; between the risk of being charged with discrimination against blacks, on the one hand, and that of being accused of discriminating against whites, on the other hand.

In an earlier article (Huysamen, in press), the psychometric properties of various models of selection fairness were presented. The purpose of the present article is to review the subsequent history of the application of these models in personnel selection in the United States and to point out the role of socio-political considerations that have impeded the wide-spread application of them. (In terms of the American situation, blacks and whites represent the lower-scoring and higher-scoring groups, respectively, referred to in that article.)

Requests for reprints should be addressed to G.K. Huysamen, Department of Psychology, The University of the Orange Free State, P.O. Box 339, Bloemfontein 9300

A distinction should be made between the psychometric soundness of these models and the socio-political acceptability of their implications. Socio-political considerations that are relevant in this connection, include civil-rights legislation and affirmative-action demands. Although the concept of affirmative action is not clearly defined, it is generally understood to imply a concerted effort to increase the appointment to jobs and the admission to college of members from previously disadvantaged groups. Basically it thus has a socio-political objective.

It will be recalled that the regression and equal-risk models of selection fairness were regarded as psychometrically acceptable procedures. However, the application of these models failed to select sufficient numbers of blacks to satisfy affirmative-action demands. Instead, it has turned out that the quota models that have some serious psychometric deficiencies (cf. Huysamen, in press), have curried more favour with affirmative-action proponents than did the psychometrically sounder regression and equal-risk models. This preference undoubtedly stems from the fact that the quota models have selected members of previously disadvantaged groups in higher proportions.

CIVIL RIGHTS LEGISLATION AND DIFFERENTIAL RATES OF SELECTION

In the introductory paragraphs, reference was made to the lower rate at which aptitude tests selected individuals from some population groups than from others, a situation referred to as adverse or disparate impact in the United States. The legal repercussions associated with this phenomenon was spurned by Title VII of the (American) Civil Rights Act of 1964 which prohibited discrimination in appointment and promotion decisions, the refusal of employment, or the differential evaluation of individuals on the basis of race, colour, religion, sex or national origin. Specifically, it made it unlawful for an employer to 'limit, segregate, or classify his employees or applicants for employment in any way which would deprive . . . any individual of employment opportunities . . . because of such individual's race, color, religion, sex, or national origin'

(Section 703a). The Tower amendment to this Act (known after the senator who introduced it) contains the only direct reference to testing in so far as it states that it shall not be deemed unlawful

> to give and to act upon the results of any professionally developed ability tests, provided that such a test, its administration or action upon the result is not designed, intended, or used to discriminate because of race, color, religion, sex, or national origin (Sackett & Wilk, 1994, p. 941).

Initial interpretation of the Act took it to imply a prohibition of intentional discrimination only. Later on interpretation centred on the quotas selected by means of tests, even if no such deliberate discrimination was intended.

Eventually most testing litigation relied on Title VII, as interpreted by the Supreme Court in *Griggs vs. Duke Power Co.* of 1971 and by federal agencies responsible for civil rights enforcement (Sackett & Wilk, 1994). The Duke Power Company in North Carolina classified employees in rank-ordered categories ranging from labourers (sweeping and cleaning) to laboratory workers, and admission at any of these levels and promotion from one level to the next were based on, among other things, scores on certain tests. These were the Wonderlic Personnel Test (a general intelligence test) and the Bennett Mechanical Comprehension Test (a mechanical aptitude test). However, each of these tests rejected a higher proportion of black than white applicants. There were 95 employees at the power station of whom 14 were black. Only one of these 14 blacks had been promoted to a level beyond that of labourer. So Willie Griggs, one of the 13 labourers, filed suit against the Duke Power Company, claiming discrimination against blacks. Because the company was unable to show that their selection system predicted job performance, the Supreme Court ruled that they were guilty of adverse impact. According to Paragraph 14(b) of the US Department of Justice guidelines, a test is said to result in adverse impact if its use causes a 'substantially different rate of selection . . . which works to the disadvantage of members of a race, sex, or ethnic group' (Lawsche, 1987, p. 492). This meant that the use of even unbiased tests were unlawful if (a) disparate impact could be shown, and (b) the employer could not demonstrate the validity of the test for the job in question. Sackett and Wilk (1994) pointed out that this requirement of job-relatedness only applied to public sector jobs that were obliged to use merit system rules. As far as employers in the private sector were concerned, job-relatedness need only to be shown if their employment practices resulted in adverse impact.

Initially, what constituted a 'substantial difference' in selection ratios for different groups, and what would be considered to be an acceptable level of job-relatedness were not specified. In their 1978 *Uniform Guidelines on Employee Selection Procedures*, the federal agencies responsible for civil rights enforcement proposed the so-called four-fifths rule which said that if the selection ratio of the black applicant group divided by the selection ratio of the white applicant group fell below 0,80, it qualified as evidence of adverse impact. Suppose that from among 20 black applicants, 12 were selected, and from among 100 white applicants, 80 were selected. Such an instance would be taken as evidence of adverse impact because the ratio of the former selection ratio to the latter selection ratio equalled 0,60/0,80 = 0,75, and this ratio fell below 0,80. In addition, these guidelines also described how job-relatedness could be demonstrated by means of each of criterion-related validity, content validity and construct validity.

Lawshe (1987), among others, showed that the concept of adverse impact was seriously flawed. He furnished data to show that the adverse impact ratio, as defined above, varied greatly between differently located plants of the same company as well as over time. Lawshe listed various factors that could result in different impact ratios from one location to the next. These ranged from the nature of the applicant pool (whether the whites were first, second or third generation Americans, whether the blacks were first, second or third generation urbanites, etc.) to the labour relations and hiring history of the company involved.

Although the verdict in *Griggs v. Duke Power Co.* initially had a seriously paralyzing effect on employment testing, it compelled personnel psychology to take test bias and criterion-related validity seriously (Gottfredson, 1994). Until the late 1970s, psy-chometricians were concerned that this verdict would require employers to carry out numerous separate validation studies to demonstrate the validity of their selection instruments for various jobs in various locations. However, Schmidt and Hunter's (1977) notion of validity generalisation and their research evidence obviated the need for such expensive, time-consuming and often practically unfeasible research. These researches proposed that if the results of studies were corrected for distortions due to sample size and range restriction, professionally developed cognitive ability tests would be found to be valid predictors of training for and performance in all jobs in all settings. They estimated the correlation between test scores and job performance to be in the vicinity of 0,50. However, hopes that the development of more valid and less biased tests would result in the elimination of diverse impact were soon dashed. Research and experience attested to the fact that 'disparate impact is the rule, not the exception, even when using valid, unbiased testes – especially cognitive tests, which are generally the best predictors of job performance' (Gottfredson, 1994, p. 955). This discrepancy in test scores for different groups thus did not reflect bias, but differences in the job-related attributes that these tests were designed to measure.

THE GATB AND DIFFERENTIAL RATES OF SELECTION

The divergent opinions regarding differential rates of selection were highlighted by the developments surrounding the application of General Ability Test Battery (GATB) results in the 1980s and early 1990s. This battery was administered by the US Employment Service, an agency which had been created in the 1930s by the United States Congress with the specific aim of promoting the absorption of disadvantaged job applicants into the economic mainstream. Registrants at the branches of this employment service indicated their ethnic-group membership on their application forms and were referred to potential employers on the basis of their GATB results. This battery included several subtests of cognitive functioning (e.g., verbal, numerical and spatial aptitude) as well as psychomotor sub-tests for motor coordination and manual and finger dexterity. Extensive research, including meta-analyses, suggested that the GATB, if used correctly, was a valid predictor of a wide range of job performances, that it was not biased against blacks or Hispanics, and that its increased use would result in greater worker productivity (Gottfredson, 1994). These positive findings, coupled with the increased business orientation of the Reagan administration, prompted the US Department of Labour to suggest that the GATB be used throughout the US Employment Service as a screening device for almost all jobs. Sackett and Wilk (1994) estimated that even if only one-third to one-half of all 19 million registrants at this employment service were tested on the GATB, it would exceed by far all testing programs in the United States in size.

However, the GATB means for blacks and Hispanics were about one standard deviation and about one-half standard deviation, respectively, lower than the mean for whites. Sackett and Wilk (1994) compiled a table that showed the percentages of Hispanics and blacks that would be referred to potential employers for the combinations of each of several such standardised group differences, on the one hand, and selection ratios of 10%, 50% and 90% for whites, on the other hand. As could be expected, for any given selection ratio for the white group, the selection ratio for the black group decreased as the standardised group difference between groups increased. Also, for any given standardised group difference, a decrease in the selection ratio for whites is accompanied by an even further decrease in the selection ratio for blacks. For

example, if 50% of the white applicants were referred on the basis of their GATB test scores, only 31% of the Hispanic applicants and a meagre 16% of the black applicants would be so referred. Furthermore, if the selection ratio for whites was 10%, only 3,8% of the Hispanics and 1,3% of the blacks would be referred. Obviously, with such discrepancies the possibility of adverse impact was unavoidable if the GATB results were to be used for hiring purposes. Moreover, research has shown that these differential rates are not exclusive to the GATB but are fairly typical of other valid employment tests (Wigdor & Garner, 1982) and tests of general cognitive ability (Schmidt, 1988). Although such tests cover a variety of conceptually distinct abilities (such as verbal, numerical and reasoning ability), they are substantially intercorrelated. This has led to the hypothesis that these correlations and the correlations of the tests with performance in a variety of jobs, are accounted for by the common *g* factor of general cognitive ability.

WITHIN-GROUP NORMING AS A POSSIBLE SOLUTION

Rather than using any of the fairness models to select individuals in terms of their GATB results, the US Employment Service opted for so-called within-group norming to prevent employers from violating the four-fifths rule of disparate impact. This practice entailed that for each of five different job families, separate percentile conversion tables were compiled for blacks, Hispanics and a combined groups of mainly whites and Asians. Reminiscent of practices in South Africa during the heyday of apartheid, each applicant's GATB score was thus converted into a percentile score with reference to his or her own racial or ethnic group. For example, for Job Family J (skilled jobs), a GATB raw score of 300 corresponded to the 79th percentile rank for blacks, the 62th percentile rank for Hispanics and the 38th percentile rank for the combined group of whites and Asians, respectively (Sackett & Wilk, 1994). Viewed from another perspective, a white or Asian person typically would have to score in the vicinity of the 84th percentile determined for the total population to have comparable chances of referral compared with a black person with a percentile rank of about 50 (Gottfredson, 1994). Thus, if an employer requested a list of all candidates scoring above, say, the 75th percentile, the US Employment Service would send him or her a list of all the white, Hispanic and black persons in the applicant pool who scored above the 75th percentile within their respective groups. This practice resulted in the numbers of referrals from these groups to be approximately proportional to the relative sizes of their respective groups in the total applicant pool. For example, if ten percent of the applicant pool was black, ten percent of the group referred to an employer would be black. The average percentile rank for the various subgroups referred would be more or less the same but the mean GATB scores for them would be substantially different. For example, a percentile rank of 75 would correspond to a substantially lower raw score on the GATB for black applicants than for white applicants.

In 1986, the Assistant Attorney General for Civil Rights charged that the GATB referral system represented intentional racial discrimination in that it advanced the interests of one group at the expense of another. Within-group norming was branded as a quota system and quotas had an unacceptable political ring to it. In 1986 the Department of Labour (via its Employment Service) and the Justice Department came to an agreement to suspend the practice of within-group norming pending the findings of a committee of the National Research Council of the National Academy of Sciences.

In the resulting report (Hartigan & Wigdor, 1989) the committee agreed that the GATB was valid for most of the purposes for which the US Employment Service used it although only modestly so. Tenopyr (1990) suggested that the lower validities obtained could have been due to the invalidity of the criteria used rather than the invalidity of the tests. Furthermore, the committee agreed that this battery was not biased against blacks and Hispanics, but actually overpredicted black job performance. However, the committee was concerned about the proportionally greater incidence of incorrect acceptances among white applicants (than among black applicants) and the proportionally greater incidence of incorrect rejections among blacks (than among whites). (These different proportions can be seen in Figure 1 in which the incorrect rejections appear in the upper left quadrant and the incorrect acceptances in the lower right quadrant. The proportions for whites are indicated by means of the dotted areas; those for the blacks by the diagonally shaded areas.) In an attempt to reconcile the competing objectives of increased black employment and enhanced workforce productivity, the committee proposed so-called performance-based adjustments. These adjustments entailed that the gap in test performance between the groups was narrowed proportionately to the invalidity of the test. The argument seemed to have been that 'the burden of (test) imperfections should not fall disproportionately on minority groups' (Sackett & Wilk, p. 933). In the simplest case, the score adjustments reduced this gap by a factor of $(1 - {}^{r}XY^{2})$. For example, if the test had a correlation of 0,80 with job performance, the 1,00 standardised mean white-black difference would be reduced by 36% (to 0,64), whereas if this correlation was only 0,20 the difference involved would be reduced by only 0,04 (to 0,96). Because most GATB validities were estimated to be falling somewhere between 0,20 and 0,40, the resulting adjustments would be 'virtually identical' to those of the previously used within-group norming. The committee expressed a 'slight preference' for the latter practice (Hartigan & Wigdor, 1989, p. 271) because of its greater simplicity.

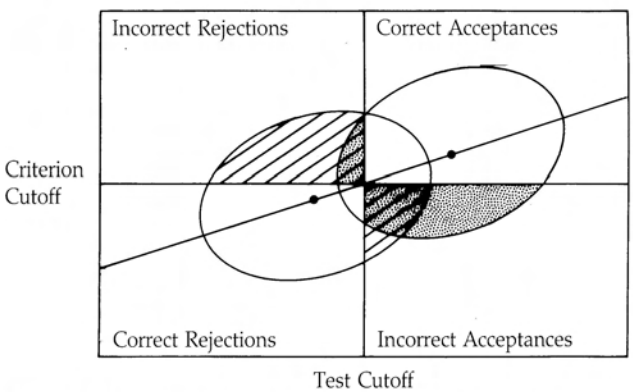

Figure 1: Outcomes resulting from test and criterion cutoff scores for groups differing in both test and criterion means

However, the basic practice of within-group norming of the GATB was attacked by several researchers (e.g., Humphreys, 1989; Blits & Gottfredson, 1990; Tenopyr, 1990), the general public and the media alike. Gottfredson (1994) pointed out that only tests that were statistically biased in favour of lower-scoring groups and perfectly valid tests would pass muster with the committee's criterion of fairness when groups differ in test performance. By definition, test fairness to blacks and Hispanics would imply unfairness against whites and Asians. With less than perfect validity, all low scorers are subject to prediction errors whereas only the scores of blacks and Hispanics are adjusted. As can be seen from Figure 1, the proportionally higher representation of blacks among the incorrect rejections and the proportionally higher incidence of whites among incorrect acceptances are due to the fact that blacks constituted the lower-scoring group and not to bias in the test.

Blits and Gottfredson (1990) called the practice of within-group norming a back-slide into medieval feudalism in which individual rights were subordinate to group rights. In terms of the latter, a person's rights resided in the group to which he or she belonged and were thus determined by birth. Because within-group norming allowed blacks to compete with other blacks but not with other whites, it would lead to the perpetuation of the very inequality it sought to eliminate.

Sackett (a member of the said committee) and Wilk (1994) emphasised that the committee intended this style of within-group norming of the GATB as a screening device only (as is reflected in the title of their article). In other words, applicants would still be referred in terms of their within-group percentiles as was done before, but it was up to potential employers to base their final decision whether or not to hire on other considerations that would be more closely related to the demands of their respective hiring organisations. According to these authors, characteristics such as punctuality, regular attendance, the ability to work in small, interactive groups, and sociability were factors not measured by the GATB but which could be of varying importance in different job situations. (Of course, a psychometrically more acceptable way to accommodate this aim would have been to compile different multiple regression equations for different groups. This practice would have weighed blacks higher on job-related predictors other than GATB on which they were considered to excel.)

The committee's concern with the high proportion of incorrect rejections among black applicants could of course be accommodated in terms of the conditional-probability model reviewed earlier (Huysamen, in press). It will be recalled that the conditional-probability model sets the proportion of successes among those selected equal for the different groups. However, this model is internally inconsistent in the sense that it yields results that are inconsistent with those of its converse counter-part, and it produces very different estimates of selection unfairness, depending on where pass-fail cutoffs are set (Gottfredson, 1994). As pointed out earlier, this model addresses the high proportion of incorrect rejections among lower-scoring groups by lowering the test cutoff score for these groups. However such a step increases the proportion of incorrect acceptances for the latter groups. Also, as pointed out before, the present model implicitly attached a greater priority to the avoidance of incorrect rejections than to the avoidance of incorrect acceptances and assigned a lower priority to the avoidance of black incorrect acceptances than to white incorrect acceptances. Such concerns are best accommodated in decision-theoretic models which require that utilities, reflective of the prevalent socio-political values, be assigned to the four pos- sible outcomes for white applicants and to those for black applicants.

The legal risks involved in using the GATB eventually led to the abandonment of this instrument by most state employment services (F.L. Schmidt, personal communication, May 25, 1995). Currently the GATB is being revised by, among other things, including some biographical questions in it. Apparently this is being done with a view to producing an instrument that does not result in differential rates of selection. However, available evidence does not seem to hold out much promise for the delivery of such a device.

RECENT LEGISLATION AND COURT RULINGS

Following the 1964 Civil Rights Act, there have been several legal developments that affected affirmative-action policies and personnel-selection practices. In 1989 the Supreme Court's verdict in *Wards Cove Packing Company vs. Antonio* provided some reprieve for employers and personnel psychologists (Schmidt, Ones & Hunter, 1992). Instead of requiring employers to demonstrate job-relatedness (as the Grigg's decision did), it required a far less stringent requirement, namely, to merely provide 'a legitimate business purpose' for the selection procedure. Moreover, consonant with other fields of civil law, it laid the burden of disproving the defendant's 'business purpose' at the plaintiff's door. However, this reprieve was shortlived. Two years after this court verdict, the Civil Rights Bill of 1991 was passed virtually unopposed by both Democrats and Republicans, despite an earlier veto by President Bush. In effect, this bill seemed to have cancelled several Supreme Court verdicts and apparently reverted the legally acceptable position of personnel selection to even more stringent demands on the employer than those laid down by the *Griggs vs. Duke Power Co.* (1971) verdict (Schmidt et al., 1992).

Moreover, Section 106 of the Civil Rights Act of 1991 prohibited within-group norming, stating that it would be unlawful to 'adjust the scores of, use different cutoffs for, or otherwise alter the results of employment related tests on the basis of race, color, religion, sex or national origin.' According to Brown (1994, p. 927), Director of Testing and Assessment of the American Psychological Association, '(t)he ramifications of the Act are more far-reaching than Congress envisioned when it considered the amendment and could mean that many personality tests and physical ability tests that rely on separate scoring for men and women are outlawed in employment selection.' Sackett and Wilk (1994) further elaborated on this point of view.

Recent Supreme Court rulings may have a profound effect on issues relating to affirmative action. In May 1995 the Supreme Court ruled that a University of Maryland scholarship programme for black students was unconstitutional. (This programme was instituted to increase black enrollment at an institution which barred blacks until as recently as 1950.) In June 1995 the Supreme Court ruled in favour of Aderand Construction of Colorado Springs (in Colorado) in *Aderand vs. Pena* (the latter being the Minister of Transport). This case arose from the awarding of a state building contract to a Hispanic company despite the fact that Aderand Construction Company submitted a substantially lower tender.

PUBLIC OPINION

Not only personnel psychologists, politicians and members of the legal professions (lawyers and civil rights advocates), but also the general public took part in the heated debate about the issue of within-group norming. It was to be expected that those groups who had to bear the brunt of affirmative action resulting from civil-rights legislation and Supreme Court rulings would sooner or later rally against these practices. According to three Gallup Polls, approximately 10% of the respondents favoured selection based on 'preferential treatment' as a remedy for past discrimination, and more than 80% opted for selection as assessed by means of ability tests (Warner & Steel, 1989). Even successful black opinion makers, businessmen and academicians have come out against affirmative action, saying that it discriminates against white Americans and is humiliating towards blacks. For example, Mr Armstrong Williams (1995, p. 8A) expressed this point of view in a recent opposing editorial of *USA Today*, a national newspaper in the United States, as follows:

> Affirmative action had merit as a temporary leg up, to help overcome a history of bigotry and discrimination against blacks, other minorities and women. As a permanent set-aside system, however, it is just a different shade of Jim Crow.

(Mr Williams is a prominent black Washington executive and host of a nationally syndicated radio talk show. Jim Crow is an expression that denotes discrimination against and suppression of blacks.)

Professor Thomas Sowell made a distinction between whether or not one is for or against the advancement of minority groups, on the one hand, and the method by which this was to be achieved. (Professor Sowell is a well-known black professor in social and political policies at the Hoover Institute at Stanford University.) Whereas the public was in support of the former, they did not necessarily agree with the various attempts to effect it. In a recent interview, he (Sowell, 1994) expressed the opinion that the black elite who had been better off in any case, benefited greatly from affirmative-action programmes, whereas those who dropped out of the system, fared poorer than not only the general population, but white dropouts as well.

According to a recent Los Angeles Times poll, 73% of its readers were of the opinion that race and gender should in no way be taken into account in the allocation of public jobs and resources (Barber, 1995). Under the leadership of two California State University professors, the California Civil Rights Initiative has been established that aims to put an end to affirmative action. This group argues that affirmative action, in terms of which minorities (blacks and Hispanics) and women are appointed in the civil service, boils down to reverse discrimination. Attempts are under way to collect 616 000 signatures on a petition to request a statewide referendum on this issue in 1996 (Robinson, 1995).

In the light of the new Republican majority in Congress, it appears safe to say that the final chapter in legislation affecting affirmative-action policies and personnel-selection practices has not been written. In the meantime, Governor Pete Wilson of California issued an order which ended dozens of programmes aimed at improving employment and education opportunities for women and ethnic minorities (Goodavage, 1995). The significance of this development cannot be ignored, because California is the most populous state in the USA and it was the first to institute an affirmative-action programme in 1964. Following Governor Wilson's action, the regents of the University of California system decided to end affirmative action in student admissions and staff appointments (Fineman, 1995).

THE MORAL OF THE STORY

In the final analysis, one's position on the extent to which affirmative action should be implemented and on the merits of within-group norming and various fairness models depends on one's socio-political values. Almost two decades ago, Cronbach (1976) clearly pointed this out in an article entitled 'Equity in selection – where psychometrics and political philosophy meet'. For example, often the same evidence is interpreted differently by proponents and opponents of affirmative action. The 1988 special issue of the *Journal of Vocational Behaviour* that was devoted to selection fairness, bears testimony to such differences of opinion. For example, one author (Seymour) disputed another's (Schmidt) meta-analytic findings that objective tests were unbiased, arguing instead that many tests were biased and that claims for validity generalisation were based on scientifically deficient procedures. In their introduction to the special issue, Gottfredson and Sharf (1988, p. 228) quoted Seymour as saying that selection tests 'can be the engine for the exclusion of racial minorities more permanent and thorough than individual ill will.' However, Seymour's position was attacked by several commentators who showed that the test bias that he purportedly revealed did not relate to race but to different levels of ability irrespective of race.

Furthermore, some commentators (e.g., Goldstein and Patterson) were of the opinion that the efforts of personnel selection were essential in bringing about racial parity. Others (e.g., Ryanen), by contrast, contended that it had 'stultified personnel selection research and led to new forms of racial discrimination and patronage in public institutions' (Gottfredson & Sharf, 1988, p. 229). Similarly, the evidence put forward to show that affirmative action would reduce worker productivity was disputed by some. Opposing groups even tend to use different terms to denote the same concepts. Thus, what has been called within-group norming by its proponents, has been referred to as race-conscious norming by its opponents; adverse impact in the former group's vocabulary, is termed disparate impact in that of the latter group. Commenting on the opposing opinions on these issues, Brown (1994) lamented the fact that personnel psychology did not speak with one voice and warned that this discipline was running the risk of being ignored.

Granted that one's position on the merits of the various fairness models may be explained by one's subscription to certain socio-political values, it follows that there is no psychometric, or any other scientific or technical basis, on which to make a choice between such models – this decision has to be made before psychometrics come into play. At the same time, there is no way of escaping these decisions or of sweeping them under the carpet. Any affirmative-action policy implicitly attaches certain utility values to the various selection outcomes for different groups, even if no formal selection model is explicitly adhered to. However, there is little chance of these utilities being made public as long as stakeholders are reluctant to put their cards on the table, or worse yet, deliberately try to obscure their real position. An even greater impediment to the articulation of these values is the attempt to pretend that one's values or political preferences have been derived by some form of scientific investigation. As Humphreys (1989, p. 14) put it:

> Affirmative action programs for certain minorities rest on value judgments, not on educational and psychological data or on statistical finagling with test scores. Value judgments should be made explicitly and openly, not camouflaged by rhetoric or statistical legerdemain.

Unfortunately, the American history of affirmative action in the area of personnel selection is rife with such rhetoric, if not plain intellectual dishonesty. For example, although affirmative action requires that some sort of preferential treatment be afforded certain previously disadvantaged groups, it is deemed discriminatory to obtain an indication of whether or not any particular individual belongs to such a group. Taylor (1987) recounted how one of his interviewees admitted that despite the psychometric merits of the regression model, it had been used infrequently because its calculation of separate regression lines for different ethnic/racial groups could be construed as being in conflict with discrimination legislation. Similarly, although quotas have become a dirty word politically, employers risk litigation if their workforce does not satisfy the four-fifths rule of differential impact. So, despite the unpalatability of identifying one's employees by race, one has no choice but to resort to some kind of bean sorting and counting, as one official has referred to this kind of bookkeeping (Taylor, 1987). To cope with such ambiguity, personnel selectors engage in euphemisms such as 'getting your numbers right', etc. in an attempt to avoid referring to quotas.

Given this double talk, it is no wonder that the decision-theoretic models have met with little success in the United States. As pointed out earlier (Huysamen, in press), these models are ideally suited to accommodating the socio-political acceptability of different outcomes for different population groups. However, to identify these groups, let alone to attach different utilities to the same outcome for different groups, may prove to be politically unacceptable.

There seems to be general agreement that some form of affirmative action is unavoidable and indeed necessary in South Africa at present. However, if the American experience is anything to go by, South African personnel psychologists will be confronted inevitably by a situation in which they have to identify the lesser of two evils. On the one hand, if they do not hire sufficient numbers of blacks and other nonwhite groups 'to reflect society at large,' they may face constant turmoil in the workplace and even legal action. On the other hand, if they agree to hire and promote blacks in the required numbers, chances are high that workforce productivity will suffer. It should be realized that with South Africa's limited resources, no organisation would be able to sustain a large number of unproductive workers for long. The choice between these options lies outside the field of psychometrics and squarely in the socio-political arena. Some may argue that productivity is of the outmost importance and that individual merit should be the sole determiner of whether someone is hired/promoted or not. Others may contend that the general societal good should override individual merit, that apartheid has bequeathed us with a legacy of racial and ethnic

discrimination that was morally so gravely wrong that strong measures, including blatant reverse discrimination in hiring and promoting decisions, are called for. To this, the opposing camp may argue that if productivity should suffer disastrously, there simply may be no benefits to go around for anyone, regardless of race or ethnicity. Moreover, this camp may argue that to correct past inequities in this manner essentially boils down to attempting to make the proverbial right out of two wrongs. However, little progress will be made in this debate unless the utilities associated with the various selection outcomes for different population groups are explicitly spelled out. Failure to do so may create a blank cheque situation which, as the logical consequence of this metaphor suggests, may lead to bankruptcy.

Although the solution to the problem of selection fairness lies outside the field of psychometrics, a clear understanding of the present problem requires a clear explication of the psychometrics involved. With Taylor and Radford's (1986) validity data being about the only available, we don't even have sufficient evidence to say with much authority whether or not our selection tests are biased or whether or not they demonstrate differential validity. By all accounts, much groundwork thus still needs to be done. Without a clear understanding of the problem involved, those who affect personnel selection policy (from legal people to personnel professionals) are hardly in a position to make informed recommendations and decisions. This article was an effort to stimulate debate and research towards this end.

REFERENCES

Barber, S. (1995, February 12). US set for race policy showdown. *Sunday Times,* p. 11.

Blits, J.H. & Gottfredson, L.S. (1990). Equality or lasting inequality? *Society, 27,* 4-11.

Brown, D.C. (1994). Subgroup norming: Legitimate testing practice or reverse discrimination. *American Psychologist, 49,* 927-928.

Cronbach, L.J. (1976). Equity in selection – where psychometrics and political philosophy meet. *Journal of Educational Measurement, 13,* 31-41.

Fineman, H. (1995, July 31). The rollback begins. *Newsweek,* p. 30.

Goodavage, N. (1995, June 1). Shot fired on affirmative action. *USA Today,* p. 3A.

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

Gottfredson, L.S. & Sharf, J.C. (1988). Foreword: Fairness in employment testing. *Journal of Vocational Behaviour, 33,* 225-230.

Hartigan, J.A. & Wigdor, A.K. (Eds.). (1989). *Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery.* Washington, DC: National Academy Press.

Humphreys, L. (1989, 7 July). Fairness in employment testing [Letter to the editor]. *Science,* p. 14.

Huysamen, G.K. (in press). The application of fair selection models in the South African context. *Journal of Industrial Psychology.*

Lawshe, C.H. (1987). Adverse impact: Is it a viable concept? *Professional Psychology: Research and Practice, 18,* 492-497.

Robinson, S. (1995, January 15). White men fight for their rights. *Sunday Times,* p. 9

Sackett, P.R. & Wilk, S.L. (1994). Within-group norming and other forms of score adjustment in pre-employment testing. *American Psychologist, 49,* 929-954.

Schmidt, F.L. (1988). The problem of group differences in abillity test scores in employment selection. *Journal of Vocational Behavior, 33,* 272-292.

Schmidt, F.L. & Hunter, J.E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62,* 529-540.

Schmidt, F.L., Ones, D.S. & Hunter, J.E. (1992). Personnel Psychology. *Annual Review of Psychology, 43,* 627-670.

Sowell, T. (1994, Fall edition). Race and culture, values and success. *Hoover Institution Newsletter,* 8-9.

Taylor, J.M. (1987). *Fair selection practices in the USA context: Implications for R.S.A.* Human Sciences Research Council: Pretoria.

Taylor, J.M. & Radford, E.J. (1986). Psychometric, testing as an unfair labour practice. *South African Journal of Psychology, 15,* 79-86.

Tenopyr, M.L. (1990). The case of fairness. *Society, 27*(3), 17-20.

Warner, R.L. & Steele, B.S. (1989). Affirmative action in times of fiscal stress and changing value priorities: A case of women in policing. *Public Personnel Management Journal, 18,* 291-309.

Wigdor, A.K. & Garner, W.R. (Eds). (1982). *Ability testing: Uses, consequences, and controversies: Part 1. Report of the committee.* Washington, D.C.: National Academy Press.

Williams, A. (1995, January 16). End this new discrimination. *U.S.A. Today,* p. 8A.