

THE DIVERSITY–VALIDITY DILEMMA: IN SEARCH OF MINIMUM ADVERSE IMPACT AND MAXIMUM UTILITY

Author:Callie Theron¹**Affiliation:**¹Department of Industrial Psychology, Stellenbosch University, South Africa**Correspondence to:**

Callie Theron

e-mail:

ccth@sun.ac.za

Postal address:

Department of Industrial Psychology, Stellenbosch University, Private Bag X1, Matieland, Stellenbosch, 7602, South Africa

Keywords:

personnel selection; adverse impact; unfair discrimination; employment equity; diversity

Dates:

Received: 03 Nov. 2008

Accepted: 30 June 2009

Published: 26 Oct. 2009

How to cite this article:Theron, C. (2009). The diversity-validity dilemma: In search of minimum adverse impact and maximum utility. *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde*, 35(1), Art. #765, 13 pages. DOI: 10.4102/sajip.v35i1.765**This article is available at:**<http://www.sajip.co.za>**Note:**

The insightful and valuable comments and suggestions for improvement to this manuscript, which were made by two anonymous reviewers, are gratefully acknowledged. The liability for the views expressed in this manuscript, however, remains solely that of the author.

© 2009. The Authors.
Licensee: OpenJournals Publishing. This work is licensed under the Creative Commons Attribution License.

ABSTRACT

Selection from diverse groups of applicants poses the formidable challenge of developing valid selection procedures that simultaneously add value, do not discriminate unfairly and which minimise adverse impact. Valid selection procedures used in a fair, non-discriminatory manner that optimises utility, however, very often result in adverse impact against members of protected groups. More often than not, the assessment techniques used for selection are blamed for this. The conventional interpretation of adverse impact results in an erroneous diagnosis of the fundamental causes of the under-representation of protected group members and, consequently, in an inappropriate treatment of the problem.

INTRODUCTION

Selection from a diverse applicant group poses a very real and formidable challenge to the field of Industrial Psychology in South Africa. Specifically, the challenge is to develop valid selection procedures that simultaneously add value, do not discriminate unfairly and which minimise adverse impact. Organisations in South Africa have a responsibility towards equity holders and society in general to efficiently combine and transform scarce factors of production into products and services with economic utility. To succeed in such an undertaking requires competent, high-performing employees. At the same time, however, organisations in South Africa are under moral, economic, political and legal pressure to diversify their workforce. Industrial Psychology is currently failing to rise to the challenge and to satisfy all three criteria simultaneously. Valid selection procedures used in a fair, non-discriminatory manner that optimises utility very often result in adverse impact against members of protected groups.

Adverse impact in personnel selection refers to the situation where a specific selection strategy affords members of a specific group a lower likelihood of selection than is afforded members of another group. Adverse impact is indicated when there is a substantial difference in the selection ratios of groups that works to the disadvantage of members belonging to a certain group (Collins & Morris, 2008; Guion, 1991; 1998). A selection ratio less than four-fifths (4/5), or 80% of the ratio of the group with the highest selection ratio would typically be regarded as providing evidence of adverse impact on any group (Collins & Morris, 2008; Huysamen, 1996; Maxwell & Arvey, 1993).

Trends from the research literature

The origin of adverse impact is generally believed to reside in the selection instruments used for personnel selection, or in differences occurring in the latent trait being assessed. As an expression of this belief, Pyburn, Ployhart and Kravitz, for example, state:

Traditional selection practice is based on identification of the knowledge, skills, abilities, and other characteristics (KSAO's) most relevant to individual job performance. The relationship between KSAO's and job performance is nearly always linear, so individuals with higher predictor scores should perform more effectively than those with lower predictor scores (Coward & Sackett, 1990). Unfortunately, many of the most predictive KSAO's (e.g., cognitive ability) and predictor methods (e.g. assessment centers) produce varying degrees of mean subgroup differences, with racioethnic minority groups usually scoring lower than majority groups (Schmitt, Clause & Pulakos, 1996). In most realistic selection situations, these subgroup differences are large enough to reduce employment opportunities for racioethnic minority groups and women.

(Pyburn, Ployhart & Kravitz, 2008, p. 145)

In terms of the above argument, the selection instruments currently in use are also to blame for the inability of selection procedures to simultaneously ensure high-performing employees and a diverse workforce. As an expression of the latter belief, Pyburn *et al.*, for example, report:

The ability of organizations to simultaneously identify high-quality candidates and establish a diverse work force can be hindered by the fact that many of the more predictive selection procedures negatively influence the pass rate of racioethnic minority group members (non-Whites) and women.

(Pyburn *et al.*, 2008, p. 144)

Maxwell and Arvey (1993) also seem to subscribe to the abovementioned point of view when they define the standardised difference in mean predictor performance between protected and non-protected groups $((\mu_{XNP} - \mu_{XE})/\sigma_X)$ as an index of adverse impact. The four-fifths rule is normally interpreted with reference to the predictor distributions (Arvey & Faley, 1988; Guion, 1991; 1998; Hough, Oswald & Ployhart, 2001; Sackett & Ellingson, 1997; Sackett & Wilk, 1994).

The belief consequently exists that selection instruments differ in terms of the adverse impact that they impose on protected groups, and thus can be graded in terms of their relative degree of adverse impact. The extremely influential and highly respected Uniform Guidelines on Employee Selection Procedures published by the Equal Employment Opportunity Commission (EEOC) endorse this position by requiring that:

Where two or more selection procedures are available which serve the user's legitimate interest in efficient and trustworthy workmanship, and which are substantially equally valid for a given purpose, the user should

use the procedure which has been demonstrated to have the lesser adverse impact.

(EEOC, 1978, p. 38297)

The conviction that adverse impact is fundamentally determined by differences in mean predictor performance results in the investigation of various strategies to reduce such subgroup differences in the mean predictor scores in an effort to increase the representation of members of protected groups without sacrificing predictive accuracy (Ployhart & Holtz, 2008; Sackett, Schmitt, Ellingson & Kabin, 2001). Ployhart and Holtz (2008) identify 16 strategies for reducing differences in mean predictor performance, which they evaluate in terms of effectiveness. The strategies include, among others, the use of valid, non-cognitive predictors (Sackett & Ellingson, 1997; Sackett *et al.*, 2001; Schmitt, Rogers, Chan, Sheppard & Jennings, 1997); the identification and removal of culturally biased items in the predictor (Humphreys, 1986; Sackett *et al.*, 2001); the use of alternative modes of presenting predictor stimuli (Chan & Schmitt, 1997; Pulakos & Schmitt, 1996; Sackett *et al.*, 2001); and the use of coaching or orientation programmes (Sackett *et al.*, 2001).

Research objective

The question is whether the adoption of such a popular stance, suggesting that adverse impact is fundamentally determined by differences in mean predictor performance, constitutes a fruitful conceptualisation of adverse impact and, more specifically, whether the various proposed remedies that were derived from it serve the best interests of the various stakeholders involved. The objective of this article is to critically reflect on the fruitfulness of the conventional stance on adverse impact and its amelioration (Hough *et al.*, 2001). More specifically, the objective of the article is to argue that the conventional interpretation of adverse impact results in an erroneous diagnosis of the fundamental causes of the under-representation of protected group members and, consequently, inappropriate treatment of the problem. Specifically, the argument tendered in the current article is that the conventional interpretation of the concept is flawed, in so far as it fails to acknowledge that selection decisions logically should be based on expected criterion performance, estimated without systematic group-related prediction error from the predictor. The objective of the present article, consequently, is to derive an analytical expression of the regression of the criterion on the predictor, which would permit a more penetrating analysis of the manner in which differences in predictor means, criterion means, validity coefficients and selection ratios affect adverse impact if criterion inferences are derived without systematic group-related prediction error from the predictor. More specifically, the objective is to quantitatively describe the manner in which the adverse impact ratio (AIR), calculated on the estimated criterion scores derived without prediction bias from predictor scores, responds to systematic changes in the difference in predictor means, criterion means, validity coefficients and selection ratios.

Review of the literature: An alternative conceptualisation of adverse impact

Organisations exist to combine and transform scarce factors of production into products or services with economic utility.¹ In order to actualise the primary objective of the organisation, a multitude of mutually coordinated activities needs to be performed, which can be categorised as a system of inter-related organisational functions. The human resource function represents one such organisational function. The human resource function justifies its inclusion in the family of organisational functions through its commitment to contribute

1. The importance of the ensuing argument lies in it constituting the framework within which the criteria/outcomes reflected in the multi-attribute utility calculations used to evaluate and compare selection procedures have to be justified. This principle not only holds for business organisations in a free market economy, but is essentially true for all organisations, if they are to survive.

towards organisational goals. The human resource function aspires to contribute towards organisational objectives through the acquisition and maintenance of a competent and motivated workforce, as well as the effective and efficient utilisation of such a workforce. The importance of human resource management flows from the basic premise that organisational success is significantly dependent on the quality of its workforce, as well as on the way in which the workforce is utilised and managed. Despite the extreme complexity of human behaviour, employee performance can, nonetheless, be explained in terms of an intricate nomological network of latent variables characterising employees and their work environment. To the extent that close-fitting explanatory structural models could be developed for the behaviour of working man, it becomes possible to derive practical human resource interventions designed to affect either employee flows or employee stocks through deductive inference (Boudreau, 1991; Milkovich & Boudreau, 1994). Interventions designed to affect employee flows attempt to change the composition of the workforce by adding, removing or reassigning employees, with the expectation that such changes will manifest in improvements in work performance. Personnel selection constitutes the primary practical human resource intervention aimed at affecting employee flow.

The objective of personnel selection is to enhance the performance of employees by controlling the flow of employees into, and upwards, in the organisation. More specifically, the objective of personnel selection is to allow only those applicants to enter the organisation who would perform satisfactorily in their designated positions. Direct information on actual job performance in the particular position can, however, never be available at the time at which the selection decision is made. Under these circumstances, and in the absence of any (relevant) information on the applicants, there is no possibility of enhancing the quality of the decision making over that which could have been obtained by chance. This seemingly innocent, but too often ignored, dilemma points to a key fact that needs to be borne in mind continually when contemplating the psychometric merits of the predictor-centred selection model. The crucial point that needs to be appreciated is that the only alternative to random decision making (other than not taking any decision at all) would be to predict expected criterion performance actuarially (or clinically) from relevant, though limited, information available at the time of the selection decision and to base the selection decision on such criterion-referenced inferences. Ideally, selection decisions should be based on criterion inferences derived clinically or mechanically from valid predictor information available at the time of the selection decision. Such a requirement implies that the focus in personnel selection is on the criterion, rather than on the predictor from which inferences about the criterion are made (Ghiselli, Campbell & Zedeck, 1981; Schmitt, 1989). This position is implicitly acknowledged by the APA-sanctioned interpretation of validity, especially predictive validity (Ellis & Blustein, 1991; Landy, 1986; Messick, 1989; Society for Industrial and Organizational Psychology, 2003). The position, moreover, underlies the generally accepted regression-based interpretations of selection fairness (Cleary, 1968; Einhorn & Bass, 1971; Huysamen, 2002). If selection decisions are not to be based on expected criterion performance, why be concerned about whether criterion inferences may be permissibly derived from predictor scores (Ellis & Blustein, 1991; Landy, 1986; Messick, 1989; Society for Industrial and Organizational Psychology, 2003) and why be concerned about whether these inferences (i.e., the criterion estimates) contain systematic group-related error that makes them systematically too low or too high? This position also seems to have been acknowledged by Aguinis and Smith (2007), when they coined the term bias-based selection errors that occur when 'biased tests are used as if they are unbiased' (Aguinis & Smith, 2007, p. 167).

It is, however, not implied thereby that the performance level of the selected cohort, in contrast to what would have resulted

under an alternative procedure, should be the sole criterion in terms of which selection procedures and their outcomes are evaluated. In distinguishing those that would perform well from those that are likely to perform less well, the selection procedure should not systematically disadvantage members of any segment of the labour market unfairly. Applicants that have the same probability of succeeding in the job should have the same probability of obtaining the job (Guion, 1998). The monetary value of the increase in performance, as affected by the selection procedure, should, moreover, exceed the investment required to effect that performance improvement to ensure that the allocation of resources is rational. In addition, in distinguishing those that would perform well from those that are likely to perform less well, the ideal would be that the selection procedure should result in proportional representation of the various gender-racioethnic segments of the labour market at all levels of the organisation. These additional criteria (of fairness, utility and adverse impact) are, however, subservient to the primary objective of enhancing employee work performance, in so far as they serve as qualifications of the primary objective. The additional criteria should neither be denied, nor should they be elevated as independent criteria in their own right. Moreover, if a selection procedure should fail to comply with the subsidiary criteria, and specifically the adverse impact criterion, this failure should not be ignored. The critical question to consider, however, is why selection procedures fail to comply with specific additional criteria. Solutions to problems generally tend to achieve greater success if they rationally and purposefully target the true causes of the problem. The critical question to consider, therefore, is why selection procedures fail to comply with the adverse impact criterion. An inappropriate conceptualisation of adverse impact would result in an inappropriate understanding of its fundamental causes and, hence, would result in inappropriate, futile remedies.

The conventional conceptualisation of adverse impact is fundamentally flawed, in that it fails to acknowledge the fact that future job performance (i.e. the criterion) forms the basis on which applicants should be evaluated in determining their assignment to an appropriate (accept or reject) treatment (Cronbach & Gleser, 1965) in personnel selection decision making. If selection decisions are based on criterion inferences derived without predictive bias from valid predictor information available at the time at which the selection decision is made, it follows that adverse impact should be conceptualised in terms of the selection ratios for the various groups that would result from selection decision making based on the rank-ordered expected criterion performance of applicants, conditional on their test performance (derived fairly, without systematic prediction bias), rather than on the selection ratios that would have resulted if selection occurred top-down on the predictor. As selection decisions ought to be based on rank-ordered expected criterion performance, the selection ratios in question should be calculated on the $E[Y|X_i;D_i]$ distribution. The question, therefore, is whether the selection ratios based on $E[Y|X_i; D_i]$, derived fairly from the predictor measures X_{iP} , differ for protected (SR_P) and non-protected (SR_{NP}) groups.³ The standardised difference between the means of the expected criterion distributions of protected and non-protected groups should therefore serve as an index of adverse impact.

Research hypothesis

The current article is aimed at showing that the ratio SR_P/SR_{NP} will necessarily be less than unity in a strict top-down selection strategy based on $E[Y|X_i; D_i]$, to the extent that $\mu_{YP} < \mu_{YNP}$. The research discussed in this article was undertaken to show that adverse impact in criterion-referenced personnel

selection cannot be avoided by the judicious choice of selection instruments (Huysamen, 1996; Schmidt & Hunter, 1981) if the criterion distributions differ significantly across groups in terms of location and dispersion – at least, not as long as the principle of strict top-down selection applies. Selection instruments can also not be graded in terms of the degree of their adverse impact. Not even a perfectly valid selection procedure used in a strict top-down manner would be able to avoid (fair) adverse impact if $\mu_{YP} < \mu_{YNP}$. If adverse impact occurs because of differences in predictor performance across groups, which cannot be justified in terms of differences in criterion performance, it would imply that the criterion inferences derived from such test scores are biased (i.e. the selection decision making is unfair, in Cleary's⁴ (1968) sense of the term). This type of unfair/discriminatory adverse impact can be avoided, however, by eliminating the systematic, group-related prediction error.

Theron (2007) attempted to illustrate the foregoing argument by analysing a fictitious data set ($N = 200$), comprising a normally distributed criterion systematically related to a normally distributed predictor. One half of the observations was obtained from members of a protected group, with the other half being obtained from members of a non-protected group. The criterion distributions of the two groups coincided perfectly, whereas the predictor distributions differed significantly in terms of location only. An illustration such as this (Theron, 2007), based on the analysis of a single data set characterised in terms of a specific set of selection parameters, although relevant, does not provide sufficiently convincing evidence in support of the argument that adverse impact in criterion-referenced personnel selection cannot be avoided by the judicious choice of selection instruments.

RESEARCH DESIGN

Research approach

To obtain more convincing evidence would require an analytical investigation of the AIR (SR_P/SR_{NP}) that results from strict top-down selection decision making based on the rank-ordered expected criterion performance of applicants, conditional on their test performance (derived fairly and without systematic prediction bias) across a large number of selection scenarios that vary systematically in terms of a spectrum of relevant selection parameters. The research reported here deviates somewhat from the conventional quantitative study, in that the data used to investigate the AIR was not obtained by administering specific (predictor and criterion) instruments to particular samples of research participants. Rather than analysing numerous actual validation study data sets, the researcher chose to generate a sample of specific data values with which he could simulate a set of specific selection scenarios that vary systematically in terms of critical selection parameters. As a consequence, the description of the research method provided below will deviate from the conventional format, in that it will not explicitly make reference to research participants and measuring instruments. The nature of the simulated data values and the manner in which they were generated are described in the following section.

In investigating the AIR, three aspects seem to be important. Firstly, the ratio needs to be calculated for the group selection ratios resulting from selection decision making, based on the rank-ordered expected criterion performance of applicants.

2. The expected criterion performance given the predictor score and group membership.

3. SR indicates the selection ratio.

4. The Cleary (1968) model of selection fairness defines fairness in terms of the absence of differences in regression slopes and/or intercepts across the subgroups comprising the applicant population (Arvey & Faley, 1988; Maxwell & Arvey, 1993). The Cleary (1968) model argues that selection decision making, based on expected criterion performance, can be considered unfair or discriminatory if the positions that members of specific groups receive, in the rank order resulting from the decision strategy, is either systematically too low or systematically too high for members of a particular group. Such imbalances in the rank order would occur if the group membership explains variance in the (unbiased) criterion, either as a main effect or in interaction with the predictors, which is not explained by the predictors, and the selection strategy fails to take group membership into account.

Secondly, the expected criterion performance of applicants should be derived from the predictor without systematic group-related prediction error. Specifically, this would mean that, if group membership significantly explained variance in the criterion not explained by the predictor, either as a main effect and/or in interaction with the predictor, this needs to be formally taken into account when the criterion estimates⁵ are derived. Thirdly, the AIR needs to be calculated for a large number of selection scenarios that vary systematically in terms of the selection parameters (i.e. overall selection ratio, validity coefficient, mean and variance of the marginal group-specific criterion and predictor distributions) that affect the selection ratios resulting for each group. Notably, a research approach should be utilised, in which all relevant selection parameters need to be simultaneously taken into account when studying the AIR. A similar sentiment seems to have been expressed by Aguinis and Smith (2007), who strongly emphasised the need to analyse the manner in which validity, predictive bias, selection errors and adverse impact are related to each other in an integrated manner.

To achieve the objective of the current research without having to study numerous actual validation study data sets, the regression of the criterion on the predictor had to be expressed in a manner that would allow the magnitude of the regression model's parameters (and especially the magnitude of the partial regression coefficients associated with the predictor and group variables) to be expressed in terms of parameters characterising the group-specific marginal criterion and predictor distributions, as well as the group-specific bivariate predictor-criterion distributions. Such expression would allow the creation of various selection scenarios, in which the parameters (i.e., σ_{YD0} , σ_{YD1} , μ_{X0} , μ_{X1} , ρ_{XY0}) are systematically varied across scenarios and groups, to infer the nature of the regression model from each scenario that is created; to estimate the criterion scores derived from predictor scores without prediction bias; and to calculate the AIR for various selection ratios.

Research method

To develop the regression equation, a number of simplifying assumptions were made. A single predictor X was assumed, though the single predictor could be a weighted composite of predictors. The single predictor was assumed to be normally distributed and linearly related to a normally distributed composite criterion measure (Y). The assumption, moreover, was that Y is an unbiased, content-valid measure of the multidimensional criterion construct η . The constitutive definition of the criterion construct was determined by the nature of the job and the strategic objectives of the organisation concerned. The assumption was also made that the criterion and predictor are observed in a population of N cases comprising members of a protected group (D = 0) and members of a non-protected group (D = 1). The two subpopulations are assumed to be equal in size ($\mu_D = 0.50$). The validity coefficient was allowed to vary across selection scenarios, but was constrained to be equal across groups.⁶ The marginal criterion

and predictor distribution of the protected and non-protected groups were assumed to be normally distributed and to have equal variances (i.e. $\sigma_{YD0}^2 = \sigma_{YD1}^2$ and $\sigma_{XD0}^2 = \sigma_{XD1}^2$), but the difference in criterion means was allowed to vary from zero to 2,5 standard deviation difference. The predictor distributions were assumed to coincide in terms of location and distribution. In addition, it was assumed that, when group membership (represented by the dummy variable D) significantly [$p < 0.05$] explained variance in the criterion that was not explained by the predictor, it would do so as a main effect only, and not in interaction with the predictor. The assumption, therefore, was that, if the regression of the criterion on the predictor for the two groups did not coincide, it would only differ in terms of intercept, and not slope.

Derivation of the regression model

Given the foregoing assumptions, the regression of the criterion on the predictor and group membership can be expressed in raw score form as Equation 1:

$$E[Y|X; D] = \alpha + \beta_1 X + \beta_2 D \quad [1]$$

According to Ghiselli, Campbell and Zedeck (1981, p. 343), the intercept can be expressed as Equation 2

$$\alpha = \mu_Y - \beta_1 \mu_X - \beta_2 \mu_D \quad [2]$$

According to Ghiselli *et al.* (1981, p. 343), the partial regression coefficients for the predictor and the group dummy variable can be expressed as Equations 3 and 4:

$$\beta_1 = \frac{\rho_{XY} - \rho_{DY} \rho_{XD}}{\sqrt{(1 - \rho_{DY}^2)(1 - \rho_{XD}^2)}} \frac{\sigma_{YC} \sqrt{1 - \rho_{DY}^2}}{\sigma_{XC} \sqrt{1 - \rho_{XD}^2}} \quad [3]$$

$$\beta_2 = \frac{\rho_{DY} - \rho_{XY} \rho_{XD}}{\sqrt{(1 - \rho_{XY}^2)(1 - \rho_{XD}^2)}} \frac{\sigma_{YC} \sqrt{1 - \rho_{XY}^2}}{\sigma_D \sqrt{1 - \rho_{XD}^2}} \quad [4]$$

In Equations 3 and 4, σ_{YC} and σ_{XC} represent the standard deviation of the criterion and predictor distributions that results when the criterion data and the predictor data of the two groups are pooled. Given the assumption of equal variance, it follows that $\sigma_{YD0} = \sigma_{YD1} = \sigma_{YC}$ when the means of the criterion distributions coincide. The same applies to the predictor distribution. When, however, the means of the criterion distributions do not coincide, the standard deviation of the combined distribution would be larger than that of the group distributions. The same, again, would have applied to the predictor distributions, if they would have been allowed to differ in terms of the mean. To be able to solve Equations 3 and 4, when only summary descriptive parameters characterising the group-specific marginal criterion and predictor distributions are available, would therefore require an expression that defines the standard deviation for the combined distribution in terms of the descriptive parameters characterising the group-specific marginal distributions. No such expression could be traced in the literature.

Equation 5 expresses the variance of the combined predictor distribution as a function of the mean and variance of the group-specific marginal predictor distributions.⁷ The derivation of

(footnote 6 continues...)

derivation of a model that describes the regression of the criterion on the predictor in terms of parameters characterising the group-specific marginal criterion and predictor distributions and the group-specific bivariate predictor-criterion distributions. The assumption, however, is not altogether unreasonable. It appears to be generally accepted, in the USA at least, that both single group validity and differential validity occur no more than could be expected by chance (Bartlett *et al.*, 1978; Schmidt & Hunter, 1981). This does not necessarily imply that a similar situation exists in South Africa. Subsequent research should, moreover, attempt to determine the effect on the adverse impact ratio if this assumption, as well as other somewhat unrealistic simplifying assumptions, were to be relaxed.

7. Even though the current article assumes that the criterion distributions coincide, subsequent studies should consider selection scenarios in which this assumption is relaxed.

5. In the USA, the remedies for unfair selection proposed by Cleary (1968), and Einhorn and Bass (1971), referred to in the current article and outlined in Theron (2007), would apparently not be allowed (Huysamen, 2002). The problem is that section 106(1) of the 1991 Civil Rights Act (cited in Guion, 1998) prohibits the adjustment of test scores on the basis of group membership. The Civil Rights Act (1991) worded the relevant section in such broad terms that it could be interpreted to mean that it is also illegal to attach different criterion-referenced interpretations to the same test score as a function of group membership. The result seems to be that selection unfairness can be evaluated, but, once detected, cannot be rectified in terms of the logic of the model that was used to detect it. Psychometrically, such a restriction seems like an internal contradiction. If legislative thinking and psychometric rationality disagree, the latter should challenge the former. The legislative constraints should not simply be passively accepted as part of the rules that govern the manner in which the employment game is played. In South Africa, paragraph 2(b) of the Employment Equity Act. (Republic of South Africa, 1998, p. 14) could be interpreted to mean that the inclusion of a group main effect and/or group x predictor interaction effect would still be permissible, provided that these effects significantly explain unique variance in the criterion not explained by the other effects included in the regression model.

6. The assumption that the validity coefficients are equal across groups clearly is somewhat contentious. The assumption is made here primarily to simplify the

Equation 5 is shown in Appendix A.

$$\sigma^2 X_c = \left[\frac{1}{n-1} \right] \{ \sigma^2 X_0 (n_0 - 1) + n_0 \mu X_0^2 + \sigma^2 X_1 (n_1 - 1) + n_1 \mu X_1^2 - \left[\frac{1}{n} \right] [n_0^2 \mu X_0^2 + n_1^2 \mu X_1^2 + 2n_0 n_1 \mu X_0 \mu X_1] \} \quad [5]$$

Equation 6, similarly, expresses the variance of the combined criterion distribution as a function of the mean and variance of the group-specific marginal criterion distributions:

$$\sigma^2 Y_c = \left[\frac{1}{n-1} \right] \{ \sigma^2 Y_0 (n_0 - 1) + n_0 \mu Y_0^2 + \sigma^2 Y_1 (n_1 - 1) + n_1 \mu Y_1^2 - \left[\frac{1}{n} \right] [n_0^2 \mu Y_0^2 + n_1^2 \mu Y_1^2 + 2n_0 n_1 \mu Y_0 \mu Y_1] \} \quad [6]$$

In Equations 3 and 4, ρ_{YD} and ρ_{XD} represent the correlation between the criterion and group membership and the correlation between the predictor and group membership. The correlations reflect the extent to which criterion and predictor performance, respectively, are related to group membership. A significant ρ_{YD} would imply that the marginal criterion distributions for the two groups differ in terms of the mean. Again the problem arises that, to be able to solve Equations 3 and 4 when only summary descriptive parameters characterising the group-specific marginal criterion and predictor distributions are available, would require expressions that define ρ_{YD} and ρ_{XD} in terms of the descriptive parameters characterising the group-specific marginal distributions. The correlation between a continuous criterion measure and a dichotomous group membership dummy variable could be calculated by means of a point biserial correlation, shown as Equation 7a (Guilford & Fruchter, 1978, p. 310). An alternative expression of the point biserial correlation between a continuous criterion measure and a dichotomous group membership dummy variable is shown as Equation 7b (Guilford & Fruchter, 1978, p. 309). In this way it becomes possible to derive ρ_{YD} , once a selection scenario has been defined in terms of the location and distribution of the group-specific marginal criterion distributions.

$$\rho_{YD} = \rho_{pbis} = \frac{\mu Y_1 - \mu Y}{\sigma Y_c} \sqrt{\frac{p}{q}} \quad [7a]$$

$$\rho_{YD} = \rho_{pbis} = \frac{\mu Y_1 - \mu Y_0}{\sigma Y_c} \sqrt{\frac{pq}{p}} \quad [7b]$$

The correlation between a continuous predictor measure and a dichotomous group membership dummy variable could be calculated in a similar manner by means of the point biserial correlation, shown as Equation 8a (Guilford & Fruchter, 1978, p. 310). An alternative expression of the point biserial correlation between a continuous predictor measure and a dichotomous group membership dummy variable is shown as Equation 8b (Guilford & Fruchter, 1978, p. 309). In this way it also becomes possible to derive ρ_{XD} , once a selection scenario has been defined in terms of the location and distribution of the group-specific marginal predictor distributions.

$$\rho_{XD} = \rho_{pbis} = \frac{\mu X_1 - \mu X}{\sigma X_c} \sqrt{\frac{p}{q}} \quad [8a]$$

$$\rho_{XD} = \rho_{pbis} = \frac{\mu X_1 - \mu X_0}{\sigma X_c} \sqrt{\frac{pq}{p}} \quad [8b]$$

Just as Equation 7a requires the value of the mean of combined criterion distribution, so does Equation 8a require the mean of the predictor distribution that results when the data for the two groups are combined. Equation 9 expresses the mean of the combined marginal criterion distribution in terms of the means of the separate, group-specific marginal criterion distributions.

$$\mu Y_c = \frac{n_0 \mu Y_0 + n_1 \mu Y_1}{n} \quad [9]$$

Equation 10 expresses the mean of the combined marginal predictor distribution in terms of the means of the separate, group-specific marginal distributions.

$$\mu X_c = \frac{n_0 \mu X_0 + n_1 \mu X_1}{n} \quad [10]$$

The expected group-specific criterion performance associated with mean group-specific predictor performance can be shown (see Appendix B) to be the mean of the group-specific marginal criterion distribution. Equation 11 expresses such a relationship for the protected group ($D = 0$).

$$E [Y | X = \mu X_0 ; D = 0] = \mu Y_0 \quad [11]$$

Equation 12 expresses the same relationship for the non-protected group ($D = 1$).

$$E [Y | X = \mu X_1 ; D = 1] = \mu Y_1 \quad [12]$$

Equations 11 and 12 imply that the group-specific estimated criterion and group-specific actual criterion distributions coincide in terms of the mean, when criterion inferences are derived without group-related prediction bias from the predictor (Cleary, 1968). The group-specific estimated criterion and group-specific actual criterion distributions, however, do not coincide in terms of dispersion when criterion inferences are derived without group-related prediction bias from the predictor, unless $E[Y|X;D]$, derived through Equation 1, correlates at unity with the criterion. More specifically, the variance of the group-specific estimated criterion distributions will be smaller than the variance of the group-specific criterion distributions. The variance of the group-specific estimated criterion distributions for the protected group ($D = 0$) results from the application of Equation 13 (see Appendix C).

$$\sigma^2 \hat{Y}_0 = \rho_{YXD}^2 \left[\frac{\sigma Y_0^2 (n_0 - 1)}{(n_0 - 2)} \right] < \sigma^2 Y_0 \quad \text{if } \rho_{YXD}^2 < 1 \quad [13]$$

The variance of the group-specific estimated criterion distributions for the non-protected group ($D = 1$) is similarly given by Equation 14.

$$\sigma^2 \hat{Y}_1 = \rho_{YXD}^2 \left[\frac{\sigma Y_1^2 (n_1 - 1)}{(n_1 - 2)} \right] < \sigma^2 Y_1 \quad \text{if } \rho_{YXD}^2 < 1 \quad [14]$$

The validity of the fair, in Cleary's (1968) sense of the term, criterion inferences derived from the predictor is given by the multiple correlation between the observed criterion performance and the expected criterion performance derived without systematic group-related prediction error from the predictor. An expression for the multiple correlation ($P\{Y, E[Y|X;D]\}$) is shown in Equation 15 (Ghiselli *et al.*, 1981, p. 344):

$$P\{Y, E [X, D]\} = \sqrt{\frac{\rho^2_{YX} + \rho^2_{YD} - 2\rho_{YX}\rho_{YD}\rho_{XD}}{(1 - \rho^2_{XD})}} \quad [15]$$

If members of the protected group typically perform lower on the criterion than members of the non-protected group, strict top-down selection based on the actual criterion scores would result in differential selection ratios. Since the group-specific expected criterion distributions coincide, in terms of the mean, with the actual group-specific criterion distributions, differential selection ratios should also result if strict top-down selection is based on $E[Y|X; D]$. Moreover, since the validity of the fair, in Cleary's (1968) sense of the term, criterion inferences derived from the predictor will be less than unity, the variance of the group-specific expected criterion distributions will be less than the variance of the group-specific observed criterion distributions. The group-specific expected criterion distributions will, therefore, contract around the mean as a negative function of the group-specific validity coefficient. The smaller ρ^2_{YXD} the more the dispersion of the group-specific expected criterion distributions around the group-specific mean will be reduced, relative to the group-specific observed

criterion distribution. The more the dispersion of the group-specific expected criterion distributions around the group-specific mean is reduced, the greater the difference in the selection ratios for the protected and non-protected groups will become, as long as the principle of strict top-down selection, based on $E[Y|X; D]$, is retained.

The Aguinis and Smith (2007) study of the reaction of the AIR to changes in validity and predictive bias (and, by implication, therefore, to differences in the group-specific marginal criterion and predictor distributions) differs from the approach followed in the current article, in that they [a] calculate the group-specific selection ratios on the predictor (rather than on the predicted criterion) distributions, and in that they [b] adhere to the 1991 amendment of the Civil Rights Act of 1964 (Guion, 1998) prohibition of deriving differential criterion inferences from predictor scores. They do, however, come extremely close to challenging the Act's stance (Aguinis & Smith, 2007) in their argument that certain instances exist in which allowing for differential criterion inferences via group-based regression equations would have served the Act's intention to promote employment equity. Despite such important differences, their results nonetheless support the conclusion derived in the current article⁸ that adverse impact will be unavoidable as long as [a] *biased-based selection errors* (Aguinis & Smith, 2007) are avoided by deriving criterion inferences from predictor scores without prediction bias; [b] the principle of strict top-down selection is adhered to; and [c] the criterion distribution of protected and non-protected groups does not coincide. Aguinis and Smith (2007) developed a computer program that can be run online (Aguinis and Smith, 2007a) to calculate the AIR that would result if a specific selection scenario is assumed in the parameter, defined in terms of: the overall predictor and criterion means and variances; the overall validity coefficient; the group-specific predictor and criterion means and variances; and the group-specific validity coefficients. The program, moreover, compares the AIR that would result from a selection scenario if the common regression equation were used to derive the criterion estimates to the AIR that would result if the appropriate moderated regression model were used to predict criterion performance. When the group-specific criterion distributions coincided in terms of location and dispersion, the Aguinis and Smith (2007) program consistently showed that, if the assumptions made in the current article applied, all valid predictors⁹ interpreted fairly, in Cleary's (1968) sense of the term, resulted in equal selection ratios, irrespective of the magnitude of the difference in predictor distributions. The Aguinis and Smith (2007) program, moreover, showed that, if the assumptions made in the present article apply when the group-specific criterion distributions do not coincide in terms of location, all valid predictors interpreted fairly, in Cleary's (1968) sense of the term, would result in differential selection ratios, irrespective of the differences in predictor distributions.

Research procedure

A data set in which specific selection parameters were systematically varied was created to empirically investigate the AIR (SR_p/SR_{np}) that results when strict top-down selection decision making is applied on the basis of the rank-ordered

8. The fact that Aguinis and Smith (2007) derive the critical predictor cut-off score from a critical criterion cut-off score via the appropriate regression model allows such a claim to be made, despite the fact that they calculated the adverse impact ratio on the group-specific predictor distributions, whereas the current study calculated the adverse impact ratio on the group-specific expected criterion distributions. Aguinis and Smith (2007), moreover, present their findings in a predictor-centred manner and, therefore, do not directly make any of the criterion-centred claims made in the present article.

9. The program would have been more user-friendly if it had derived the total population parameters from the chosen group's specific parameters, in so far as the former depended on the latter. Moreover, the total population validity coefficient value would depend on whether differences in the regression of the criterion on the predictor across groups were explicitly acknowledged. To the extent that the regression of the criterion on the predictor would differ across groups in terms of intercept and/or slope, the validity of the selection procedure (i.e., $R[Y, E[Y|X; D]]$) would be underestimated if the difference were ignored in deriving the criterion estimates.

expected criterion performance of protected, and non-protected, group applicants, conditional on their test performance (derived fairly, without systematic prediction bias) as a function of specific selection parameters. The selection parameters that were systematically varied were [a] the difference in the means of the group-specific marginal criterion distributions; [b] the correlation between the predictor and the criterion; and [c] the selection ratio.

Each case in the simulated data set represents a selection scenario. Each selection scenario was defined in terms of the values of a set of selection parameters. The selection parameters that defined a specific selection scenario were the mean and variance of the group-specific marginal criterion and predictor distributions ($\mu_{x0}, \mu_{x1}, \mu_{y0}, \mu_{y1}, \sigma_{x0}^2, \sigma_{x1}^2, \sigma_{y0}^2, \sigma_{y1}^2$); the size of the protected and non-protected groups (n_0 and n_1); the correlation between the predictor and the criterion (ρ_{xy}); and the critical criterion cut-off (Y_k). In all the selection scenarios, the group-specific marginal predictor distributions were assumed to coincide (i.e. $\mu_{x0} = \mu_{x1}$ and $\sigma_{x0}^2 = \sigma_{x1}^2$).¹⁰ In all selection scenarios, the variance of the criterion distributions was assumed to be equal. However, the means of the group-specific marginal criterion distributions were systematically made to differ across selection scenarios in increments of 0.1 standard deviation units up to a maximum difference of 2,5 standard deviation units. When the means of the group-specific marginal criterion distributions differed, the non-protected group was assumed to perform at a higher level than did the protected group. The variance of the combined criterion distribution was subsequently derived by solving Equation 6 for the chosen values for the mean and variances of the group-specific marginal criterion distributions in each selection scenario. Likewise, the variance of the combined predictor distribution was subsequently derived by solving Equation 5 for the chosen values for the mean and variances of the group-specific marginal criterion distributions in each selection scenario. Due to the assumption that the group-specific marginal predictor distributions coincide in all the selection scenarios, $\sigma_{x0}^2 = \sigma_{x1}^2 = \sigma_{xc}^2$ in all selection scenarios. From the calculated σ_{xc}^2 and σ_{yc}^2 values for each selection scenario and the group-specific predictor and criterion means that applied to each selection scenario, ρ_{xD} and ρ_{yD} were calculated by means of Equation 7a and Equation 8a. The availability of these two correlation coefficients allowed for the calculation of the regression model parameters (α, β_1, β_2) in Equation 1 for each selection scenario by means of solving Equations 2, 3 and 4. From Equation 1, the expected criterion performance ($E[Y|X = \mu_{x0}, D = 0]$ and $E[Y|X = \mu_{x1}, D = 1]$) was calculated for each group for each selection scenario using Equation 1, conditional on the predictor being equal to the group-specific predictor mean. The multiple correlation $P[Y, E[Y|X; D]]$ was calculated for each selection scenario using Equation 15. The availability of the multiple correlation allowed for the variance of the group-specific estimated criterion distributions to be calculated for each selection scenario, using Equations 13 and 14.

Statistical analysis

The group-specific expected criterion distributions have been shown to coincide with the actual group-specific criterion distributions (Equations 11 and 12, and Appendix B) in terms of the mean, but the variance of the group-specific expected criterion distributions will be less than the variance of the group-specific observed criterion distributions (Equations 13 and 14, and Appendix C). A series of critical criterion cut-off scores (Y_k) was subsequently defined for each selection scenario. The scores were defined in terms of the number of standard deviation units by which they are positioned above or below the protected group criterion mean ($\mu_{y0} = 0; \sigma_{y0}^2 = 1$). Critical criterion cut-off values varied from 2,5 standard deviation units above the protected group criterion mean to 2,5 standard

10. Such an assumption should be relaxed in subsequent studies.

deviation units below the protected group criterion mean in steps of 0.1 standard deviation units. The relative position of the critical criterion cut-off scores in the expected criterion distribution of the protected group was then described by expressing the cut-off scores as z-scores (Z_{0,Y_k}) in the expected criterion distribution of the protected group with Equation 16:

$$Z_{0,Y_k} = \frac{Y_k - E[Y|X = \mu_{X_0}; D = 0]}{\sigma_{Y_0}} \quad [16]$$

The transformation of the relative position of Y_k in the expected criterion distribution of the protected group to a standard score allowed the selection ratio of the protected group (SR_{0,Y_k}) to be calculated for each critical cut-off score by integrating the standard normal distribution function for Y_{0v} as shown in Equation 17:

$$SR_{0,Y_k} = P(Y_0 \geq Y_k) = \int_{Z_{0,Y_k}}^{\infty} f(Y_0) dY_0 \quad [17]$$

Since the criterion mean of the non-protected group is also expressed in terms of the number of standard deviation units by which it falls above the protected group criterion mean (i.e. $\mu_{Y_{0v}}, Y_k$ and $\mu_{Y_{1v}}$ are all expressed on the same scale), the position of the chosen critical criterion cut-off scores in the expected criterion distribution of the non-protected group could be described by expressing the cut-off scores as z-scores in the expected criterion distribution of the non-protected group, with Equation 18:

$$Z_{1,Y_k} = \frac{Y_k - E[Y|X = \mu_{X_1}; D = 1]}{\sigma_{Y_1}} \quad [18]$$

The transformation of the relative position of Y_k in the expected criterion distribution of the non-protected group to a standard score, in turn, allowed the non-protected group's selection ratio (SR_{1,Y_k}) to be calculated for each critical cut-off score by integrating the standard normal distribution function for Y_{1v} as is shown in Equation 19:

$$SR_{1,Y_k} = P(Y_1 \geq Y_k) = \int_{Z_{1,Y_k}}^{\infty} f(Y_1) dY_1 \quad [19]$$

The AIR that would result from the implementation of each critical criterion cut-off score in each selection scenario was then calculated with Equation 20:

$$AIR = \frac{P(Y_0 \geq Y_k)}{P(Y_1 \geq Y_k)} = \frac{SR_{0,Y_k}}{SR_{1,Y_k}} \quad [20]$$

RESULTS

The reaction of the AIR to changes in the critical criterion cut-off score (i.e. the selection ratio) and the difference in the criterion means were then plotted graphically for specific values of the predictor validity coefficient (ρ_{XY}). Figure 1 portrays the manner in which the AIR reacts to a lowering in the critical criterion cut-off and an increase in the difference in the group-specific criterion means, when a predictor that correlates 0.30 with the criterion is used to select all applicants with $E[Y|X,D] \geq Y_k$. Figure 1, therefore, displays the extent to which the selection ratio for the protected group differs from that of the non-protected group (expressed as the ratio SR_{0v}/SR_{1v}), if all applicants with predicted criterion scores ($E[Y|X,D]$) equal to or greater than a specific criterion cut-off score (Y_k) were to be selected. Figure 1, moreover, displays how the difference in the selection ratios would change if the criterion cut-off score were to be lowered. Lowering of the criterion cut-off score would mean that the number of standard deviation units by which the cut-off falls above the protected group mean would decrease towards zero and eventually become negative (a lowering of Y_k therefore corresponds to a movement to the right on the abscissa in Figure 1). A high negative standardised cut-off score would mean that practically all applicants are selected. Figure 1 also

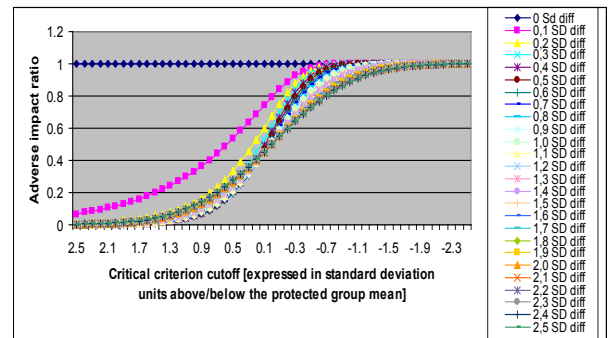


FIGURE 1
Adverse impact ratio as a function of the criterion cut-off in the protected group predicted criterion distribution and the difference in the mean predicted criterion performance: expressed in standard deviation units [$\rho_{XY} = 0,30$]

displays how the difference in the selection ratios changes if the predicted criterion distributions, which initially coincided, are gradually pulled apart.

When the predicted criterion distributions coincide, the selection ratio for the protected and non-protected groups remains the same, irrespective of the position of Y_k . However, the situation changes as soon as the predicted criterion no longer coincides in terms of the mean. Figure 1, for example, shows that, if the criterion is predicted by means of a predictor with a validity of 0.30 and the mean of the predicted criterion scores of the non-protected group is 0.1 standard deviations higher than the mean predicted criterion scores of the protected group (i.e. the pink line), and the criterion cut-off score is set to fall 2.5 standard deviation units above the protected group's mean (i.e. a small proportion of applicants is selected), then the selection ratios for the non-protected group are markedly higher than that of the protected group. When the critical cut-off score is lowered and larger proportions of applicants are selected from each group, the difference in selection ratios decreases non-linearly. Only when Y_k reaches a value that falls just below the mean of the protected group does the AIR reach the critical value of 0.80. At very low Y_k values, where practically all applicants are selected, the selection ratios essentially become the same.

Figure 2, in a similar manner, maps the way in which the value of the AIR responds when the relative position of the criterion cut-off score in the protected group's criterion distribution is gradually lowered, when a predictor that correlates 0.50 with the criterion is used to select all applicants with $E[Y|X,D] \geq Y_k$. Figure 2 portrays how the effect that the change in the critical criterion cut-off score has on the AIR changes when the criterion distributions for the protected and non-protected groups gradually migrate apart in terms of the mean. Figure 3 and Figure 4 portray the behaviour of the AIR with regard to changes in the value of Y_k and the difference in the criterion means when predictors with validity of 0.70 and 0.90, respectively, are used to select all applicants with $E[Y|X,D] \geq Y_k$.

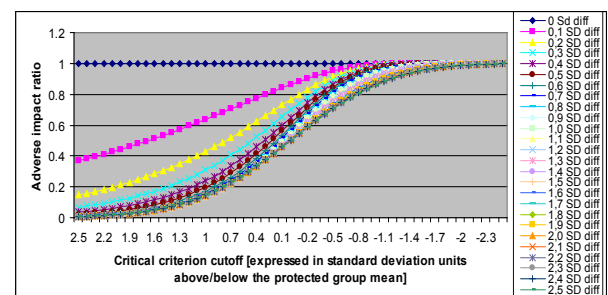


FIGURE 2
Adverse impact ratio as a function of the criterion cut-off in the protected group predicted criterion distribution and the difference in the mean predicted criterion performance: expressed in standard deviation units [$\rho_{XY} = 0,50$]

Inspection of Figures 1 to 4 indicates the following:

- All valid selection procedures used fairly, in Cleary's (1968) sense of the term, produce an AIR equal to unity, irrespective of the size of the selection ratio for the protected group when the criterion distributions coincide in terms of the mean and variance
- At a fixed validity coefficient and a fixed difference in the criterion means, the AIR decreases non-linearly as the selection ratios for the protected and non-protected groups increase [i.e. as the critical criterion cut-off value decreases]
- At a fixed difference in the criterion means and a fixed critical criterion cut-off value [i.e. the selection ratio for the protected and non-protected groups are fixed, but not equal (unless $\mu_{y0} = \mu_{y1}$)], the AIR increases with an increase in the validity of the selection predictor¹¹
- At a fixed critical criterion cut-off value [i.e. the selection ratios for the protected and non-protected groups are fixed, but not equal (unless $\mu_{y0} = \mu_{y1}$)], the AIR increases with a decrease in the difference in the criterion means
- The extent to which the AIR increases when the difference in the criterion means decreases is increased when the protected group's selection ratio decreases [i.e. as the critical criterion cut-off value increases].

The effect of the magnitude of the correlation between the predictor and the criterion is further examined in Figures 5 to 7. Such examination takes the form of plotting changes in the AIR to changes in the relative position of the critical criterion cut-off in the protected group's criterion distribution, given a fixed difference in the criterion distribution means of the protected and non-protected groups, for different predictor-criterion correlations. Figure 5, therefore, displays the extent to which the selection ratio for the protected group differs from that of the non-protected group (expressed as a ratio SR_{0y}/SR_{1y}), if all applicants with predicted criterion scores ($E[Y|X,D]$) equal to or greater than a specific criterion cut-off score (Y_k) are selected, together with how this difference is affected by a change in the value of Y_k and a change in the validity coefficient.

Inspection of Figures 5 to 7 indicates that:

- The relationship between the AIR and the relative position of the critical criterion cut-off in the protected group distribution is curvilinear
- The slope of the curvilinear relationship between the AIR and the relative position of the critical criterion cut-off in the protected group distribution decreases as the correlation between the predictor and the criterion increases
- At low protected group selection ratios, the AIR increases as the validity coefficient increases
- At high-protected group selection ratios, the AIR increases as the validity coefficient decreases
- The AIR increases as the critical criterion cut-off value is lowered in the protected group criterion distribution [i.e. the selection ratios are increased in both groups]
- The rate at which the AIR increases with a lowering of the critical criterion cut-off in the protected group criterion distribution decreases at the inflection point of the curve as the predictor-criterion correlation increases.

DISCUSSION

The objective of the current article was to derive an analytical expression of the regression of the criterion on the predictor that would permit a penetrating analysis of the manner in which differences in predictor means, criterion means, validity coefficients and selection ratios affect adverse impact if criterion inferences are derived without systematic group-related prediction error from the predictor. More specifically,

11. As the criterion distributions move apart, the group main effect explains an increasing amount of variance in the criterion that is not explained by the predictor. At a given validity coefficient, therefore, the validity of the fair selection procedure will increase as the difference in the criterion means increases.

the objective of the article was to quantitatively describe the manner in which the AIR, calculated on the estimated criterion scores derived without prediction bias from predictor scores, responds to systematic changes in the difference in predictor means, criterion means, validity coefficients and selection ratios.

In South Africa, systematic group-related differences in criterion distributions could be expected to exist as a legacy of the apartheid regime, which systematically denied members of previously disadvantaged groups the opportunity to develop the personal attributes or job competency potential required to succeed on the criterion in question. To the extent that such is, indeed, the case, the foregoing results would suggest that all valid predictors used fairly, in Cleary's (1968) sense of the term, would create adverse impact against members of previously disadvantaged groups. Under conditions where systematic group-related differences in criterion distributions exist, any attempt to alleviate the adverse impact problem by searching for alternative predictors would be futile. Achieving zero adverse impact under such conditions in strict top-down, performance-maximising selection with unbiased criterion inferences derived from valid predictors would be tantamount to psychometric alchemy. Adverse impact can be alleviated (but not eliminated) by increasing the predictive validity of the selection procedure and by increasing the selection ratio. The improvement in the AIR brought about by the increase in the selection ratio is counterproductive, however, because such an improvement is, in effect, brought about by decreasing the selection effectiveness of the selection procedure concerned.

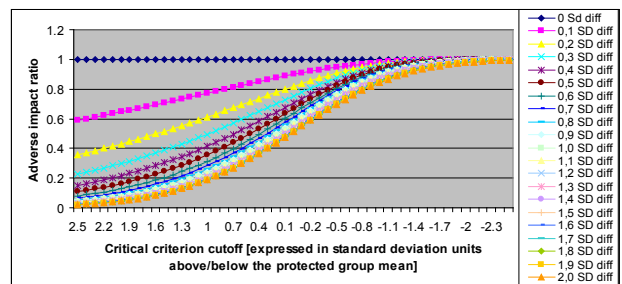


FIGURE 3

Adverse impact ratio as a function of the criterion cut-off in the protected group predicted criterion distribution and the difference in the mean predicted criterion performance: expressed in standard deviation units [$\rho_{xy} = 0,70$]

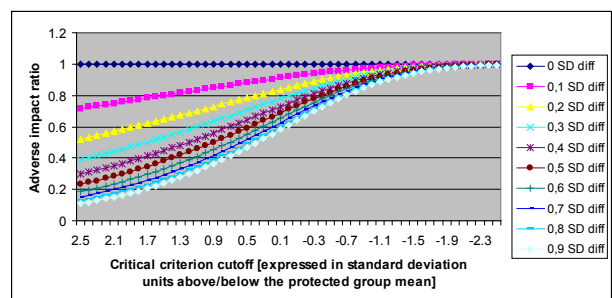


FIGURE 4

Adverse impact ratio as a function of the criterion cut-off in the protected group's predicted criterion distribution and the difference in the mean predicted criterion performance: expressed in standard deviation units [$\rho_{xy} = 0,90$]¹²

12. Increasing the difference in the criterion means between groups increases the criterion variance explained by group membership. Mathematically, this could result in multiple correlations for the fair regression model exceeding unity if at a given predictor-criterion correlation the group means would be allowed to migrate too far apart. In the case of $\rho_{XY} = 0,90$, criterion means that differ one standard deviation or more would mathematically imply an impossible scenario and hence these are not reflected in Figure 4.

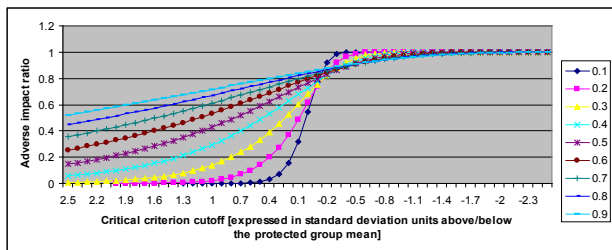


FIGURE 5
Adverse impact ratio as a function of the selection ratio and validity coefficient (difference in mean predicted criterion performance: 0,2 standard deviation units)

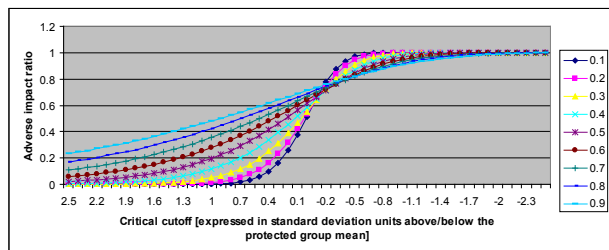


FIGURE 6
Adverse impact ratio as a function of the selection ratio and validity coefficient (difference in mean predicted criterion performance: 0,5 standard deviation units)

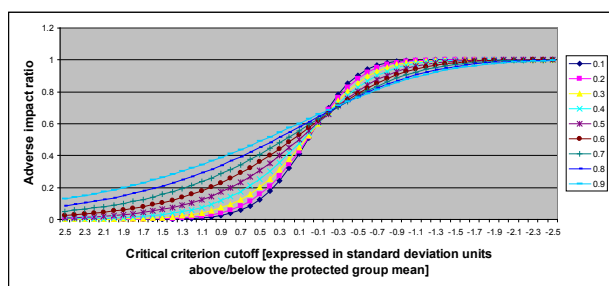


FIGURE 7
Adverse impact ratio as a function of the selection ratio and validity coefficient (difference in mean predicted criterion performance: 0,8 standard deviation units)

The essence of this finding should also apply if predictor information were to be combined clinically or subjectively (Grove & Meehl, 1996; Gatewood & Feild, 1994) in a valid and fair, in Cleary's (1968) sense of the term, manner. To the extent that the clinically derived criterion inferences are valid, as well as to the extent that they are derived without predictive bias, essentially the same mechanism would operate as in the case of mechanically derived criterion inferences. If the criterion inferences are valid, the difference in the means of the marginal group-specific distributions of clinically derived criterion inferences should reflect the difference in the means of the marginal group-specific observed criterion distributions (possibly expressed in a different metric), but the variance in the distributions of clinically derived criterion inferences would be smaller than in the observed criterion distributions (when scaled in the same metric), to the extent that the predictive validity of the clinical criterion inferences would be less than unity. Whether the clinical mind can take account of group differences in predictor and criterion distributions in a manner that would allow for the derivation of criterion inferences without predictive bias is, however, debatable (Theron, 2007).

The foregoing argument does not imply that such adverse impact should be passively accepted as unavoidable collateral damage created by the performance-maximising fair use of valid predictors in selection, though. The ideal is, and always should be, that the selection procedure should result in proportional representation of the various gender-racioethnic segments of the labour market at all levels of the organisation. When such

an ideal is not achieved, active measures should be taken to reduce the adverse impact caused by selection procedures. Such measures should be aimed at rectifying the root causes of the problem. The fact that adverse impact is created during the personnel selection process should not be construed as evidence that the selection procedure is responsible for the adverse impact. As selection decisions are based on criterion inferences derived from predictors, the fundamental cause of the adverse impact created by the performance-maximising fair use of valid predictors in selection in South Africa is the difference in the means of the criterion distributions of protected and non-protected groups. Protected group members perform systematically lower on the criterion, due to systematic, group-related differences in job competency potential latent variables required to succeed in the job, which, in turn, probably arise from systematic differences in access to development opportunities. In the South African context, searching for alternative selection instruments would be a tragically inappropriate response to the problem of adverse impact. Intellectually honest solutions to adverse impact in South Africa lie in aggressive affirmative development aimed at developing the job competency potential latent variables required to succeed in the job. With individuals differing in the extent to which they would benefit from development opportunities, given the limited training resources that should be utilised optimally, the assessment of learning potential should play a pivotal role in identifying those disadvantaged individuals whose selection would result in the most favourable return on investment in terms of affirmative development. As long as critical person characteristics that determine job performance remain underdeveloped in the protected group due to lack of opportunity, the phenomenon of adverse impact will remain a reality.

It could be argued that the foregoing argument interprets selection utility in an unnecessarily, even undesirably, narrow fashion. In terms of this argument, the value of the outcomes of selection decision making should not be judged solely in terms of the financial value of the performance of the selected cohort. The question is whether workforce diversity should not be valued as a desirable, valued outcome as well. Workforce diversity is valuable in part, because it fosters growth, innovation and progress and, in terms of the latter, the performance of organisational units. Individual human diversity should, moreover, also be valued simply for its own sake. Diversity, in such a sense, however, means much more than the superficial gender-racioethnic differences on which employment equity debates typically focus. Cascio and Aguinis argue that:

...multiattribute utility analysis (Roth & Bobko, 1997) can be a better tool to assess a selection system's usefulness to an organization. A multiattribute utility analysis includes not only the Brogden-Cronbach-Gleser result, but also information on other desired outcomes such as increased diversity, cost reduction in minority recruitment, organizational flexibility, and an organization's public image. Thus, a multiattribute utility analysis incorporates the traditional single attribute utility estimate, but goes beyond this and also considers key strategic business variables at the group and organizational levels.

(Cascio & Aguinis 2005, p. 338)

That selection procedures should be evaluated in terms of a basket of evidence seems to be in accordance with the point raised earlier in the current article, that the performance level of the selected cohort, in contrast to what would have resulted under an alternative procedure, should not be the sole criterion in terms of which selection procedures and their outcomes are evaluated. Earlier in the current article, however, it was also argued that such additional evaluation criteria should be regarded as subservient to the primary objective of enhancing employee work performance, in so far as they serve as qualifications of the primary selection objective of filtering in the best performing applicants. Such a stance probably represents an expression of a specific set of values that would not be shared by all interested parties. If it is, indeed, the case

that significant differences exist in the manner in which the different evaluation criteria are structured and/or in the value attached to the criteria, the differences should be made explicit, accepted as an integral part of the managerial landscape and continuously debated and negotiated. Such a debate would, however, require that industrial psychologists develop a clear, coherent, well-motivated and unapologetic stance on what they wish to achieve through personnel selection procedures. Developing such a stance will, however, not change current practices in and by itself. The convictions on what ideally should be achieved through personnel selection should find expression in persuasive talk and compelling actions that will demonstrate the merit of the stance taken.

The danger exists that the multi-attribute utility analysis line of reasoning could be used to adapt the Brogden-Cronbach-Gleser utility equations (Boudreau, 1991) to show that a deviation from strict top-down selection that increases workforce diversity results in a recalculated utility on par with the traditional, more narrowly interpreted utility of strict top-down selection. Such a line of reasoning seems problematic for at least two reasons. Firstly, diversity involves much more than superficial gender-ratioethnic differences. Differences in values, perceptions, beliefs, ideals and an almost infinite array of personal characteristics is what really matters. Diversity in such fundamental variables, however, cuts across gender-ratioethnic differences. Arguing that an increase in the AIR would necessarily bring about an increase in diversity with regard to those characteristics that are required to promote creative turbulence in organisations, therefore, seems questionable.

To justify a deviation from strict top-down selection in terms of such an argument in South Africa is problematic, secondly, because it seems to be rooted in an erroneous diagnosis of the problem. As argued above, it essentially treats the symptoms of the problem, rather than its fundamental underlying causes. Such justification, moreover, seems to suggest an extremely pessimistic, negative prognosis concerning affirmative development interventions, aimed at developing the job competency potential latent variables required to succeed in a particular job. The affirmative development argument optimistically believes that performance and diversity are not inherently incompatible. A sacrifice in utility (narrowly interpreted in monetary scaled performance) is not a necessary, unavoidable solution for achieving workforce diversity. In fact, the practice of affirmative development should eventually result in the multi-attribute utility equation returning figures for strict top-down selection that exceed the values currently encountered under both a narrow Brogden-Cronbach-Gleser interpretation of utility and the multi-attribute interpretation of utility in respect of preferential hiring.

Theron (2007) pleaded that practitioners involved in personnel selection should move beyond the popular rhetoric on the use of psychological tests in personnel selection and engage in an open (Louw, 1965), honest and psychometrically sophisticated, penetrating debate on the interplay between past injustices, measurement bias, selection fairness, adverse impact and selection utility. Theron (2007) conceded, however, that open, honest and psychometrically penetrating debate, in and by itself, will not achieve the extremely laudable vision that former president Mandela expressed in the preamble to the Employment Equity Bill,

that those who have been denied access to qualifications in the past can become qualified now, and that those who have been qualified all along but overlooked because of past discrimination, are at last given their due.

(Republic of South Africa, 1996, p. 5)

Concrete actions are required to translate the insights emerging from the debate into the type of selection practices that honestly serve such an inspiring vision.

The argument presented above implies an empirical, actuarial approach to practical psychological assessment. It seems unlikely that a clinical selection strategy could be adapted in a manner that would eliminate systematic prediction errors, should they be identified (Theron, 2007). The inability to adapt clinical selection strategies in a manner that would eliminate systematic prediction errors undeniably poses severe practical, technical and logistical challenges to the industrial organisational psychologist. If, however, there is some psychometric merit in the argument outlined above, the I/O Psychology fraternity needs to rise to the challenge of finding creative and innovative solutions to the obstacles that currently prevent the widespread implementation of an actuarial approach to personnel selection (Mossholder & Arvey, 1984). The problem of small sample sizes (and the concomitant lack of statistical power), as well as that of standardisation in criterion conceptualisation and measurement, require attention. The development of generic performance structural models that map an inter-related network of competency potential latent variables (Saville & Holdsworth, 2000; 2001) on to an inter-related network of competency latent variables (Saville & Holdsworth, 2000; 2001), and that, in turn, map the latter on to an inter-related network of outcome latent variables, could provide a feasible solution to such problems. The existence of such generic (consisting of managerial, technical, sales, and administrative) performance structural models would foster standardisation and would permit validation studies to be performed across numerous specific small N settings.

Statistical power is a matter of particular concern, due to the nature of the statistical analyses required to ensure valid, fair, utility-maximising selection. Cleary's (1968) interpretation of selection fairness plays a pivotal role in the argument presented in the current article. Moderated regression (Bartlett *et al.*, 1978; Berenson, Levine & Goldstein, 1983; Lautenschlager & Mendoza, 1986) is typically used to evaluate whether group membership significantly explains variance in the criterion when included in a regression model (as a group main effect and/or as a group x predictor interaction effect) that already includes the predictor. The evaluation of predictive bias by means of moderated multiple regression analysis is plagued, however, by statistical power problems (Aguinis, 1995; Aguinis & Stone-Romero, 1997; Aguinis, Beaty, Boik & Pierce, 2005) that increase the risk of not rejecting the null hypotheses that the group main effect and, especially the group x predictor interaction effect, explain variance in the criterion in a model that already contains the predictor ($H_0: \beta_2 = \beta_3 = 0 | \beta_1 \neq 0$), when, in fact, the regression equations do not coincide in the parameter. The accuracy of the criterion inferences derived from predictor scores (i.e. the validity coefficient $\rho[Y, E[Y|X]]$) will decrease if the nature of the relationship between criterion and predictor is not accurately understood. The inaccurate modelling of the criterion-predictor relationship is, therefore, quite correctly included as biased-based selection error in the integrative framework proposed by Aguinis and Smith (2007). Including biased-based selection error in the analytical framework, however, only serves to clearly underscore the consequences of an incorrect decision being made regarding the nature of the criterion-predictor relationship. For the individual practitioner/decision maker, the problem remains whether the specific moderated regression analysis he/she performed on his/her validation sample resulted in the appropriate decision. The development of generic performance structural models again seems to offer at least a partial solution, in so far as it would permit large-scale (i.e. large N) validation studies to be performed across numerous specific small N settings. The burden of performing the required validation studies could thereby be collectively shouldered by the fraternity of industrial psychologists, instead of by the individual organisation.

Although adverse impact is created during the process of personnel selection, the predictors are not the villains that should be blamed for the adverse impact. The solution to the

problem should, therefore, not be sought there. A search for alternative selection instruments will not solve the problem of adverse impact. The fundamental cause of the adverse impact in South Africa lies in underdeveloped job competency potential latent variables, which are required to succeed in the job. The intellectually honest solution to adverse impact in South Africa, therefore, lies in aggressive affirmative development aimed at developing the job competency potential latent variables key to such success.

APPENDIX A

The variance of the predictor distribution of the combined population can be expressed as Equation A1:

$$\begin{aligned}\sigma^2 X_c &= \left(\frac{1}{n-1}\right) \left[\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\right) \left(\sum_{i=1}^n X_i\right)^2 \right] \\ &= \left(\frac{1}{n-1}\right) \left[\sum_{i=1}^{n_0} X_{0i}^2 + \sum_{i=1}^{n_1} X_{1i}^2 - \left(\frac{1}{n}\right) \left(\sum_{i=1}^{n_0} X_{0i} + \sum_{i=1}^{n_1} X_{1i}\right)^2 \right] \\ &= \left(\frac{1}{n-1}\right) \left\{ \sum_{i=1}^{n_0} X_{0i}^2 + \sum_{i=1}^{n_1} X_{1i}^2 - \left(\frac{1}{n}\right) \left[\left(\sum_{i=1}^{n_0} X_{0i}\right)^2 + \left(\sum_{i=1}^{n_1} X_{1i}\right)^2 + \dots \right. \right. \\ &\quad \left. \left. 2 \sum_{i=1}^{n_0} X_{0i} \sum_{i=1}^{n_1} X_{1i} \right] \right\} \\ &= \left(\frac{1}{n-1}\right) \left\{ \sigma^2 X_0 (n_0 - 1) + n_0 \mu^2 X_0 + \sigma^2 X_1 (n_1 - 1) + n_1 \mu^2 X_1 - \left(\frac{1}{n}\right) \right. \\ &\quad \left. [n^2 \mu X_0 + n^2 \mu X_1 + 2n_0 n_1 \mu X_0 \mu X_1] \right\} \quad [A1]\end{aligned}$$

The sum of the squared raw predictor scores within the protected group is given by Equation A2:

$$\begin{aligned}\sigma^2 X_0 &= \left(\frac{1}{n_0 - 1}\right) \left[\sum_{i=1}^{n_0} X_{0i}^2 - \left(\frac{1}{n_0}\right) \left(\sum_{i=1}^{n_0} X_{0i}\right)^2 \right] \\ \therefore \sigma^2 X_0 (n_0 - 1) &= \sum_{i=1}^{n_0} X_{0i}^2 - \left(\frac{1}{n_0}\right) \left(\sum_{i=1}^{n_0} X_{0i}\right)^2 \\ \therefore \sigma^2 X_0 (n_0 - 1) + \left(\frac{1}{n_0}\right) \left(\sum_{i=1}^{n_0} X_{0i}\right)^2 &= \sum_{i=1}^{n_0} X_{0i}^2 \\ \therefore \sum_{i=1}^{n_0} X_{0i}^2 &= \sigma^2 X_0 (n_0 - 1) + \left(\frac{1}{n_0}\right) (n_0 \mu^2 X_0) \\ \therefore \sum_{i=1}^{n_0} X_{0i}^2 &= \sigma^2 X_0 (n_0 - 1) + n_0 \mu^2 X_0 \quad [A2]\end{aligned}$$

The sum of the raw predictor scores within the protected group squared is given by Equation A3:

$$\left(\sum_{i=1}^{n_0} X_{0i}\right)^2 = n_0 \mu^2 X_0 \quad [A3]$$

Substituting Equation A2 and Equation A3 in Equation A1 results in Equation A4, which expresses the variance of the predictor distribution of the combined population as a function of the sample size, mean and variance of the separate subpopulation predictor distributions:

$$\begin{aligned}\sigma^2 X_c &= \left[\frac{1}{n-1}\right] \left\{ \sigma^2 X_0 (n_0 - 1) + n_0 \mu^2 X_0 + \sigma^2 X_1 (n_1 - 1) + n_1 \mu^2 X_1 - \left[\frac{1}{n}\right] \right. \\ &\quad \left. [n^2 \mu X_0 + n^2 \mu X_1 + 2n_0 n_1 \mu X_0 \mu X_1] \right\} \quad [A4]\end{aligned}$$

Similarly, the variance of the criterion distribution of the combined population can be expressed as a function of the sample size, mean and variance of the separate subpopulation criterion distributions by means of Equation A5:

$$\begin{aligned}\sigma^2 Y_c &= \left[\frac{1}{n-1}\right] \left\{ \sigma^2 Y_0 (n_0 - 1) + n_0 \mu^2 Y_0^2 + \sigma^2 Y_1 (n_1 - 1) + n_1 \mu^2 Y_1^2 - \left[\frac{1}{n}\right] \right. \\ &\quad \left. [n^2 \mu Y_0 + n^2 \mu Y_1 + 2n_0 n_1 \mu Y_0 \mu Y_1] \right\} \quad [A5]\end{aligned}$$

APPENDIX B

If a valid predictor, linearly related to the criterion, would be assumed, as well as that the group membership main effect, significantly [$p < 0.05$], explains variance in the criterion not explained by the predictor, the regression of the criterion on the predictor and group membership could, in raw score form, be expressed as Equation B1:

$$E[Y|X; D] = \alpha + \beta_1 X + \beta_2 D \quad [B1]$$

The expected criterion performance associated with the mean predictor performance of the protected group ($D = 0$) can therefore be expressed as Equation B2:

$$\begin{aligned}E[Y|X = \mu_{X_0}; D = 0] &= \alpha + \beta_1 \mu_{X_0} \\ &= \mu Y - \beta_1 \mu X - \beta_2 \mu D + \beta_1 \mu X_0 \quad [B2]\end{aligned}$$

If it is assumed that the predictor distributions coincide, it follows that $\mu_{X_0} = \mu_{X_1} = \mu_X$, so, therefore, Equation 2 can be expressed as Equation B3:

$$\begin{aligned}E[Y|X = \mu_X; D = 0] &= \mu Y + \beta_1 \mu X - \beta_2 \mu D + \beta_1 \mu X \\ &= \mu Y - \beta_1 (\mu X - \mu D) + \beta_1 \mu D \\ &= \mu Y - \beta_2 \mu D \\ &= \mu Y - 0, 5\beta_2 \quad [B3]\end{aligned}$$

The expected criterion performance associated with the mean predictor performance of the non-protected group ($D = 1$) can be expressed as Equation B4, if it is assumed that $\mu_{X_0} = \mu_{X_1} = \mu_X$:

$$\begin{aligned}E[Y|X = \mu_X; D = 1] &= \mu Y + \beta_1 \mu X + \beta_2 D \\ &= \mu Y - \beta_1 \mu X - \beta_2 \mu D + \beta_1 \mu D + \beta_2 \\ &= \mu Y - \beta_1 (\mu X - \mu D) - 0, 5\beta_2 + \beta_2 \\ &= \mu Y + 0, 5\beta_2 \quad [B4]\end{aligned}$$

According to Equation 4, the partial regression coefficient in Equation B1 for the group dummy variable is given by (Ghiselli *et al.*, 1981, p. 343):

$$\beta_2 = \frac{\rho_{DY} - \rho_{XY} \rho_{XD}}{\sqrt{(1 - \rho_{XY}^2)(1 - \rho_{XD}^2)}} \frac{\sigma_{YC} \sqrt{1 - \rho_{XY}^2}}{\sigma_D \sqrt{1 - \rho_{XD}^2}} \quad [B5]$$

If it is assumed that the predictor distributions of the two groups coincide, it follows that $\rho_{XD} = 0$. Equation B5 can be simplified and expressed as Equation B6:

$$\begin{aligned}\beta_2 &= \frac{\rho_{DY}}{\sqrt{(1 - \rho_{XY}^2)}} \frac{\sigma_{YC} \sqrt{1 - \rho_{XY}^2}}{\sigma_D} \\ &= \frac{\rho_{DY} \sigma_{YC}}{\sigma_D} \quad [B6]\end{aligned}$$

According to Equation 7b (Guilford & Fruchter, 1978, p. 309), the correlation between the group dummy variable and the criterion can be expressed as B7:

$$\rho_{YD} = \rho_{pbis} = \frac{\mu Y_1 - \mu Y_0}{\sigma_{Y_c}} \sqrt{pq} \quad [B7]$$

Substituting Equation B7 into Equation B6 results in Equation B8.

$$\beta_2 = \frac{\mu Y_1 - \mu Y_0}{\sigma_{Y_c}} \sqrt{pq} \frac{\sigma_{Y_c}}{\sigma_D} \quad [B8]$$

Since $s_D = \sqrt{pq}$, Equation B8 can be simplified to Equation B9:

$$\beta_2 = \mu_{Y_1} - \mu_{Y_2} \quad [B9]$$

Substituting Equation B9 into Equation B3 results in Equation B10 given that the groups are assumed to be of equal size:

$$\begin{aligned} E[Y|X = \mu_{X_0}, D=0] &= \mu_Y - 0.5\beta_2 \\ &= \mu_Y - \left(\frac{\mu_{Y_1} - \mu_{Y_2}}{2}\right) \\ &= \mu_Y - \left(\frac{\mu_{Y_0} - 2\mu_{Y_0} + \mu_{Y_1}}{2}\right) \\ &= \mu_Y - \left[\left(\frac{\mu_{Y_0} + \mu_{Y_1}}{2}\right) - \left(\frac{2\mu_{Y_0}}{2}\right)\right] \\ &= \mu_Y - \mu_Y + \mu_{Y_0} \\ &= \mu_{Y_0} \end{aligned} \quad [B10]$$

Substituting Equation B9 into Equation B4 results in Equation B11 given that the groups are assumed to be of equal size:

$$\begin{aligned} E[Y|X = \mu_{X_1}, D=1] &= \mu_Y + 0.5\beta_2 \\ &= \mu_Y + \left(\frac{\mu_{Y_1} - \mu_{Y_2}}{2}\right) \\ &= \mu_Y + \left(\frac{2\mu_{Y_1} - \mu_{Y_1} + \mu_{Y_0}}{2}\right) \\ &= \mu_Y + \left[\left(\frac{2\mu_{Y_1}}{2}\right) - \left(\frac{\mu_{Y_1} + \mu_{Y_0}}{2}\right)\right] \\ &= \mu_Y - \mu_Y + \mu_{Y_1} \\ &= \mu_{Y_1} \end{aligned} \quad [B11]$$

APPENDIX C

The squared multiple correlation between the weighted linear combination of the predictor [X] and the group main effect [D] and the criterion can be expressed as Equation C1:

$$P^2_{Y\hat{Y}} = \frac{SS_{\hat{Y}_c}}{SS_{Y_c}} \quad [C1]$$

The sum of squared deviations of the predicted criterion scores from the criterion mean can, therefore, through cross-multiplication, be expressed as Equation C2:

$$SS_{\hat{Y}_c} = P^2_{Y\hat{Y}} SS_{Y_c} \quad [C2]$$

The variance of the observed criterion scores of the combined criterion distribution can be expressed as Equation C3:

$$\sigma^2_{Y_c} = \frac{SS_{Y_c}}{n-1} \quad [C3]$$

The sum of squared deviations of the criterion scores from the criterion mean can, therefore, be expressed as Equation C4:

$$SS_{Y_c} = \sigma^2_{Y_c} (n-1) \quad [C4]$$

The variance of the predicted criterion scores of the combined predicted criterion distribution can be expressed as Equation C5:

$$\sigma^2_{\hat{Y}_c} = \frac{SS_{\hat{Y}_c}}{n-2} \quad [C5]$$

Substituting Equation C2 and Equation C4 into Equation C5 results in an expression of the variance of the predicted criterion scores expressed as a function of the criterion variance and the squared multiple correlation, shown as Equation C6:

$$\sigma^2_{\hat{Y}_c} = \frac{P^2_{Y\hat{Y}}[\sigma^2_{Y_c}(n-1)]}{n-2} \quad [C6]$$

REFERENCES

- Aguinis, H. (1995). Statistical power problems with moderated regression in management research. *Journal of Management*, 21, 1141-1158.
- Aguinis, H., Beaty, J.C., Boik, R.J., & Pierce, C.A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30 year review. *Journal of Applied Psychology*, 90, 94-107.
- Aguinis, H., & Smith, M.A. (2007). Understanding the impact of test validity and bias on selection errors and adverse impact in human resource selection. *Personnel Psychology*, 60, 165-199.
- Aguinis, H., & Smith, M.A. (2007a). Selection programme. Retrieved July 15, 2009, from <http://www.cudenver.edu/~hag>
- Aguinis, H., & Stone-Romero, E.F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology*, 82, 192-206.
- Arvey, R.D., & Faley, R.H. (1988). *Fairness in selecting employees* (2nd edn.). Reading: Addison-Wesley.
- Bartlett, C.J., Bobko, P., Mosier, S.B., & Hannan, R. (1978). Testing for fairness with a moderated multiple regression strategy: An alternative to differential analysis. *Personnel Psychology*, 31, 233-242.
- Berenson, M.L., Levine, D.M., & Goldstein, M. (1983). *Intermediate statistical methods and applications*. Englewood Cliffs: Prentice-Hall.
- Boudreau, J.W. (1991). Utility analysis for decisions in human resource management. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd edn., Vol. 2, pp. 621-745). Palo Alto: Consulting Psychologists Press.
- Cascio, W.F., & Aguinis, H. (2005). *Applied psychology in human resource management* (6th edn.). Upper Saddle River: Prentice Hall.
- Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgement tests: Sub-group differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82, 143-159.
- Cleary, T.A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, 5, 115-124.
- Collins, M.W., & Morris, S.B. (2008). Testing for adverse impact when sample size is small. *Journal of Applied Psychology*, 93(2), 463-471.
- Coward W.M., & Sackett P.R. (1990). Linearity of ability-performance relationships: a reconfirmation. *Journal of Applied Psychology*, 75, 297-300.
- Cronbach, L.J., & Gleser, G.C. (1965). *Psychological tests and personnel decisions*. (2nd edn.). Urbana: University of Illinois Press.
- Einhorn, H.J., & Bass, A.R. (1971). Methodological considerations relevant to discrimination in employment testing. *Psychological Bulletin*, 75, 261-269.
- Ellis, M.V., & Blustein, D.L. (1991). Developing and using educational and psychological tests and measures: The unificationist perspective. *Journal of Counseling and Development*, 69, 550-555.
- Equal Employment Opportunity Commission (EEOC). (1978). Uniform Guidelines on Employee Selection Procedures. *Federal Register*, 35(149), 12333-12336.

- Gatewood, R.B., & Feild, H.S. (1994). *Human resource selection*. (3rd edn.). Fort Worth: Dryden.
- Ghiselli, E.E., Campbell, J.P., & Zedeck, S. (1981). *Measurement theory for the behavioural sciences*. San Francisco: Freeman.
- Grove, W.M., & Meehl, P.E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2, 293–323.
- Guilford, J.P., & Fruchter, B. (1978). *Fundamental statistics in psychology and education*. Tokyo: McGraw-Hill.
- Guion, R.M. (1991). Personnel assessment, selection and placement. In M.D. Dunnette & L.M. Hough (Eds.). *Handbook of industrial and organizational psychology* (2nd edn., vol. 2, pp. 327–397). Palo Alto: Consulting Psychologists Press.
- Guion, R.M. (1998). *Assessment, measurement and prediction for personnel decisions*. Mahwah: Lawrence Erlbaum.
- Hough, L.M., Oswald, F.L., & Ployhart, R.E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues evidence and lessons learned. *International Journal of Selection and Assessment*, 9, 152–194.
- Humphreys, L.G. (1986). An analysis and evaluation of test and item bias in the prediction context. *Journal of Applied Psychology*, 71, 327–333.
- Huysamen, G.K. (1996). The socio-political context of the application of fair selection models in the USA. *Journal of Industrial Psychology*, 22, 1–6.
- Huysamen, G.K. (2002). The relevance of the new APA standards for educational and psychological testing for employment testing in South Africa. *South African Journal of Psychology*, 32, 26–33.
- Landy, F.J. (1986). Stamp collecting versus science: Validation as hypothesis testing. *American Psychologist*, 41, 1183–1192.
- Lautenschlager, G.J., & Mendoza, J.L. (1986). A step-down hierarchical multiple regression analysis for examining hypotheses about test bias in prediction. *Applied Psychological Measurement*, 10, 133–139.
- Louw, N.P. van Wyk (1965). Die oop gesprek. In E. Botha (Ed.). *Afrikaanse essayiste*. Cape Town: Human & Rousseau.
- Maxwell, S.E., & Arvey, R.D. (1993). The search for predictors with high validity and low adverse impact: Compatible or incompatible goals? *Journal of Applied Psychology*, 78, 433–437.
- Messick, S. (1989). *Validity*. In R.L. Linn (Ed.). *Educational measurement*. (3rd edn.). New York: American Council on Education and McMillan.
- Milkovich, G.T., & Boudreau, J.W. (1994). *Human resource management*. (7th edn.). Homewood: Richard D. Irwin.
- Mossholder, K.W., & Arvey, R.D. (1984). Synthetic validity: A conceptual and comparative review. *Journal of Applied Psychology*, 69, 322–333.
- Ployhart, R.E., & Holtz, B.C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61, 153–172.
- Pulakos, E.D., & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance*, 9, 241–258.
- Pyburn, K.M., Ployhart, R.E., & Kravitz, D.A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology*, 61, 143–151.
- Republic of South Africa (1996). Employment Equity Bill. *Government Gazette*, No. 18481, 1 December 1996.
- Republic of South Africa. (1998). Employment Equity Act. *Government Gazette*, No. 19370, 19 October 1998.
- Roth, P.L., & Bobko, P. (1997). A research agenda for multi-attribute utility analysis in human resource management. *Human Management Review*, 7(3), 341–368.
- Sackett, P.R., & Ellingson, J.E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology*, 50, 707–721.
- Sackett, P.R., Schmitt, N., Ellingson, J.E., & Kabin, M.B. (2001). High stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist*, 56, 302–318.
- Sackett, P.R., & Wilk, S.L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist*, 49, 929–954.
- Saville & Holdsworth (2000). Competency design: Towards an integrated human resource management system. *SHL Newslines*, March, 7–8.
- Saville & Holdsworth (2001). Competencies and performance@ work. *SHL Newslines*, May, 6.
- Schmidt, F.L., & Hunter, J.E. (1981). Employment testing; Old theories and new research findings. *American Psychologist*, 36, 1128–1137.
- Schmitt, N. (1989). Fairness in employment selection. In M. Smith & I. Robertson (Eds.). *Advances in selection and assessment*. Chichester: John Wiley.
- Schmitt N, Clause C.S., & Pulakos E.D. (1996). Subgroup differences associated with different measures of some job-relevant constructs. In C.R. Cooper & I.T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology* (Volume 11, pp. 115–140). New York: Wiley.
- Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology*, 82, 719–730.
- Smith, M., & Robertson, I. (1989). *Advances in selection and assessment*. Chichester: John Wiley. Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures*. Bowling Green: Society for Industrial and Organizational Psychology.
- Theron, C.C. (2007). Confessions, scapegoats and flying pigs: Psychometric testing and the law. *SA Journal of Industrial Psychology*, 33 (1), 102–117.