

CULTURE AS MODERATOR VARIABLE IN PSYCHOLOGICAL TEST PERFORMANCE: ISSUES AND TRENDS IN SOUTH AFRICA

B. BEDELL
R. VAN EEDEN
F. VAN STADEN

*Department of Psychology
University of South Africa*

ABSTRACT

This article deals with the recognition of cultural variation and related variables as moderators of psychological test performance in South Africa. Attention is paid to historical approaches, present issues and trends emerging in this field. The studies discussed include those on cognitive and personality tests and factors surrounding their use and interpretation. The comparability of test scores and how this relates to bias and fairness are discussed. Related perspectives from industry and questions as to future options regarding assessment are also raised.

OPSOMMING

Hierdie artikel handel oor die erkenning van die moderator- of bemiddelingsrol wat kulturele variasie en verwante veranderlikes speel by sielkundige toetsprestasie in Suid Afrika. Aandag word geskenk aan historiese benaderings, huidige kwessies en neigings wat in die veld ontstaan. Studies van kognitiewe en persoonlikheidstoets en faktore wat hierdie toetse se gebruik en interpretasie omvou, word bespreek. Die verband tussen die vergelykbaarheid van toetstellings en sydigheid en billikheid word aangedui. Verwante sienings vanuit die nywerheids-wêreld, en vrae oor toekomstige moontlikhede met betrekking tot evaluering word ook gelug.

Psychological tests are not necessarily equally accurate or successful for all subgroups in a heterogeneous population. However, testing provides a means of gaining information about people in a fairly quick, economic, and objective way. When used in context together with other information, test results are also seen to enhance decision making. Despite the resistance currently experienced, psychological tests therefore are and will probably continue to be widely applied in South Africa (Foxcroft, 1997b; Nell, 1994). There is, however, increasing pressure on test developers and test users to ensure fair testing practices (e.g. the Employment Equity Bill; the draft policy of the Professional Board of Psychology on the Classification of Psychometric Measuring Devices, Instruments, Methods and Techniques). According to legislation only psychological tests and similar instruments of which the validity and reliability have been scientifically proven and that are not biased against any employee or group may be used. Although Foxcroft (1997b) supports the adaptation of existing tests and the development of culturally appropriate tests and norms, she also refers to the difficulties in developing and norming tests in a culturally and linguistically diverse society. The issue of the comparability of test performance has been investigated in research on tests developed for specific racial groups as well as common tests developed for more than one racial group (Claassen, 1997). Racial groups are culturally relatively homogeneous but tests developers have to address the fact that individuals and groups are in various stages of acculturation to a western society (Shuttleworth-Jordan, 1996). There are furthermore other variables such as educational and socio-economic background that are also important moderators of test performance. This has implications for the delineation of norm groups and where the target population is heterogeneous with regard to one or more of these variables, also for the interpretation of test results.

The current article reviews several South African studies in the field of cross-cultural psychological testing. This includes findings related to the functioning of specific cognitive and personality tests across various groups. The effect of moderator variables on the results of these tests and efforts to address this issue are also discussed. It is furthermore indicated how cultural recognition led to certain trends in test development and

test use such as the measurement of potential and increasing sensitivity towards the context in which the individual functions. Although this article focusses on psychometric tests and testing, the issues related to bias and fairness discussed here are also relevant for other forms of assessment.

The need for change has developed gradually throughout the history of psychometric testing in Africa and South Africa. Around the beginning of this century tests were imported from abroad and applied *mutatis mutandis* on all sectors of the community. From 1920 studies and assessments of black people became somewhat more systematic, and empirically orientated. There was a dawning notion of the inappropriateness of instruments used and the trend was toward rejecting standard measurements (Irvine & Berry, 1988). Through the 1940s and 50s work in South Africa in the psychometric domain was focused rather pragmatically on the educability and trainability of black South Africans. It was recognised that cultural differences influence testing outcomes, and attempts to create "culture-free" tests were in vogue. Following this period, development of tests of adaptability dominated interest. There was no attempt to assess cognitive competence in a culturally relevant framework (Kendall, Verster & Von Mollendorff, 1988). However, in the 70s and 80s there was growing recognition of the fact that culture exerts subtle and pervasive effects in the testing domain, and that it is not possible to remove culture from the equation (Retief, 1988). Culture affects behaviour and consequently the psychological constructs being measured and it was beginning to be seen as an important moderator (Baron & Kenny, 1986) of test performance. A move towards the consideration of cultural bias inherent in the tests themselves furthermore strengthened the notion that culture may constitute a source of systematic error in test results.

Kendall et al. (1988) state that test scores correlate with non-test variables such as test-taking behaviour, cultural and/or environmental factors, and dispositional factors. Test-taking behaviour concerns such factors as level of education, language, and practice or familiarity with tests and testing. Cultural and environmental factors relate to aspects such as the effects of rural or urban residence, acculturation, and various preferences (and their effects) associated with ethnicity. Dispositional factors include those of age and gender, both of which correlate with test performance (as they do elsewhere in the world).

Van der Flier and Drenth (in Taylor & Radford, 1986) note that the main issue in the analysis of test score comparability is whether the same construct or dimension is being measured across different groups. This means that the researcher or test user must establish whether intergroup differences on a test reflect real differences in the construct measured or to what extent these differences are artefacts of the testing situation or of other factors pertaining to the test. In other words the researcher must establish whether the test is biased. Poortinga and Van der Flier (1988) identified the following as important questions in this regard:

- What traits or abilities are valid across cultures?
- Does a certain construct play an equally important role in the organisation of individuals from different cultures?
- Does a test for assessing a particular construct reflect that ability on a scale with the same origin and with equal units of measurement in the different groups?

Comparability of test performance across groups can be divided into three main categories, namely construct, score, and predictive comparability (Owen, 1996). Construct comparability essentially concerns what is being measured and if the same variables are measured for different groups. Score comparability relates to item bias. If items are biased against members of a group, it means that they have the ability to respond to the item correctly, but for reasons such as item format, or phrasing of the item, they do not have the same opportunity as others to do so. Predictive comparability concerns the extent to which the test predicts performance on a criterion equally well for different groups; under- or over-prediction on the criterion is systematically related to group membership.

Bias concerns systematic errors associated with group membership. Taylor and Radford (1986) point out that sources of bias may lie within test items, within the test as a whole, within subjects, with the tester, and in the testing context. They argue that bias can never be eliminated entirely, but steps can be taken to minimise the effects of bias associated with known or potential sources. Variables commonly identified as moderating test performance include language proficiency, cultural, socio-economic and educational background, and test-wisness (Nell, 1997).

COGNITIVE TESTING IN A MULTICULTURAL CONTEXT

To date, in South Africa research on multicultural testing in personology and abnormal or clinical psychology has received relatively little attention compared to cognitive testing. Most cognitive tests used in the country assess verbal ability, numerical ability, deductive reasoning, and the like. These can to a large extent be viewed as indicators of crystallized intelligence (following Cattell's theory) and often concerns specialized skills or knowledge promoted or required by a given culture (Taylor, 1994). This structural approach attempts to measure performance along dimensions assumed to constitute the fundamental structure of domains such as cognition.

Research in the field of cross-cultural testing includes studies and discussions on the following cognitive tests: The General Scholastic Aptitude Test (*GSAT*) (Claassen, 1990; Claassen & Schepers, 1990; Hugo & Claassen, 1991), the New South African Group Test (*NSAGT*) (Claassen & Cudeck, 1985), Raven's Standard Progressive Matrices (*SPM*) (Owen, 1992), the Senior South African Individual Scale-Revised (*SSAIS-R*) (Van Eeden, 1993; Van Eeden & Van Tonder, 1995; Van Eeden & Visser, 1992), the Paper and Pencil Games (*PPG*) (Claassen, 1996), the Bender Gestalt (*BG*) (Viljoen, Levett, Tredoux & Anderson, 1994), the South African Wechsler-Bellevue Adult Intelligence Scale (*SAWAIS* or *SAW-B*) (Nell, 1994; Shuttleworth-Jordan, 1995), the Junior Aptitude Tests (*JAT*) (Owen, 1989, 1991, 1992), the Mental Alertness Test and the High Level Figure Classification Test (Holburn, 1992), and the Junior South African Individual Scales (*JSAIS*) (Van den Berg, 1985).

For both the *SPM* (a non-verbal test) and the *JAT* (Owen, 1989, 1991, 1992) similarities were found in the constructs being measured for various groups (white, coloured, black and Indian). Item bias could however not be excluded and large differences were found in mean score performance between groups. Owen studied the content and format of items that functioned differentially and suggested reasons for bias. Language (especially in the case of the black subjects who were tested in English) and cognitive style (subject-related) were identified as possible sources of bias. Training in language and problem-solving strategies were recommended as the differences in mean test performance preclude the use of common norms, while the use of separate norms for the different population groups defeats the purpose of having a common test (Owen, 1992). Other studies have shown that tests, even though they contain biased items, are nevertheless considered among the best available measures for their purpose. Such is the case with the High Level Figure Classification Test (Holburn, 1992). It should be noted that the type of cognitive tests discussed here are usually norm-referenced.

The *SSAIS-R* and the *GSAT* were developed jointly for whites, coloureds and Indians. Although the *SSAIS-R* can be regarded as reliable and valid for the different groups, some variations in factor structure and with regard to predictive validity were found and should be taken into consideration when making comparisons across groups (Van Eeden & Visser, 1992). Environmental deprivation was furthermore shown to explain part of the variance in performance on this test and separate norms were consequently developed for non-environmentally disadvantaged children and for a norm group including all Afrikaans-speaking and English-speaking children. Similarly Claassen and Schepers (1990) found that differences in *GSAT* scores between population groups could to an extent be explained in terms of environmental variables associated with different levels of socio-economic status. Claassen (1990) also found that while the *GSAT* had similar factor structures for different population groups, factorial invariance across the groups was not demonstrated. This means that while the test broadly measured the same constructs in the various population groups studied, the abilities measured did not share the same pattern of relationships in the different population groups. Claassen partly contributes this to the fact that the *GSAT* to some extent measures the culmination of previous learning experiences. In South Africa, with its history of differential educational opportunities based on race, this needs to be kept in mind when interpreting the results of the *GSAT* and other tests like it.

Performance on the *SSAIS-R* requires a certain proficiency in English, and different population groups vary in their English proficiency partially as a result of differential educational opportunities (Van Eeden, 1993; Van Eeden & Van Tonder, 1995). Although a number of tests have been translated into African languages (Van den Berg, 1996) there are practical problems with this procedure such as the large number of official languages and the availability of test administrators who can speak the respective languages. Practitioners furthermore report problems with regard to different dialects spoken in different areas and a difference in performance between urban and rural individuals tested in their mother tongue. Foxcroft (1997b) also refers to the fact that many black pupils are educated in their mother tongue as well as in English. The decision whether to translate a test or not should be based on information about the effect of respondents' understanding of the language of testing on their performance. This information also has implications for the interpretation of test results. Considering the effect of English proficiency on test results is consequently one of the main concerns in the current South African standardization of the Wechsler Adult Intelligence Scale III (*WAIS III*).

Research findings suggest that while intra-group comparison of subjects within a particular teaching or training situation is often justifiable, caution and sensitivity should be used in cross-group comparisons. The nature and extent of group dif-

ferences are particularly relevant when the test behaviour of groups is compared and when inferences are made about situations other than that of the test, based on test behaviour. When predicting from scores to some criterion, it is at times difficult to know if one has obtained a true reflection or measure of a person's intellect, or if the factors mentioned have somehow skewed the results (Hugo & Claassen, 1991). If the measures are systematically biased, such bias is expected to result in systematic prediction errors. The test user can consequently make provision for these errors.

In neuropsychological testing research on local use of overseas tests is limited, especially with regard to cross-cultural issues. Nevertheless, these tests are frequently administered in South Africa as part of neuropsychological assessment and clinicians make inferences using normative data that is based on foreign populations. Tollman and Msengana (1990) discuss problems in interpreting test results if culture, language and education are not taken into account. Their study deals with an adaptation of Luria's Neuropsychological Investigation to accommodate Zulu speakers. In the study by Viljoen et al. (1994) Zulu speakers were found to perform quite differently on the Bender Gestalt when compared to Americans. The unconditional use of foreign norms is therefore not encouraged, since these might not be appropriate for South Africa.

Shuttleworth-Jordan (1996) argues in favour of adapting and standardizing well-researched international tests. Given the heterogeneity of the South African society and the facts of increasing urbanization, acculturation, and literacy, she foresees problems with the development of culturally appropriate measures. Indeed, Verster and Prinsloo's (1988) finding that the gap in test performance between subpopulations is correlated with cultural convergence and exposure to a common formal educational curriculum lends credence to this argument. According to Nell (1997) there are groups in South Africa, for example illiterate individuals, for whom this would not be a viable option. Based on cross-cultural experience and empirical results he has recommended a battery of psychological and neuropsychological tests to be used for persons with less than 12 years of education. Instructions for this battery is available in a number of African languages (Nell, 1999). Foxcroft (1997a) states that the concept of the nature of intelligence and thereby the scope of intelligence tests should be expanded. Kendall et al. (1988, p.321) also point out that "when it is considered that investigators throughout the world have concentrated on measuring only those abilities which Western . . . societies deem important, it is dangerous to conclude that they constitute universally valid yardsticks of intelligence per se". Thus, what intelligence tests – especially those in the structural tradition – measure would appear to be open to cultural variation, and it is debatable if these tests can be considered, in a multicultural context, to give an indication of general intelligence. In this vein, Bakare (in Kendall et al., 1988) states that it is perhaps improper to use Western tests to measure the 'intelligence' of people in Africa. Such tests can measure people's spatial ability, or perceptual speed, or visual-motor co-ordination, and whatever other cognitive components such tests measure, but it is erroneous to assume that such elements are the sum total of intelligence. Indeed, Ferguson (1956) argued long before this that culture dictates what will be learned, by whom, and at what age. Thus, culture plays a role in the development and shaping of the pattern of abilities manifested by members of different cultural groups.

As an alternative to traditional tests, Foxcroft (1997a) suggests an approach to assessment that reflects theoretical developments on cognitive functioning. In her presentation at a symposium on the possible standardization of the *WAIS III*, she discussed the planning-attention-simultaneous-successive (PASS) model of cognitive functioning as an example of a process-oriented approach that questions the notion of the general intelligence construct. According to this model, effective processing requires the integration of the knowledge base of the individual and the processes underlying a particular task (planning, attention, etc.). Foxcroft argues that the traditional ap-

proach to measuring cognitive functioning is limited and that assessment based on a process-oriented approach provides the opportunity for theory-based intervention. One of her recommendations was that this approach be used in combination with a test such as the *WAIS III*. Research will however be needed to guide this type of decision as little has been done with regard to information-processing tests in terms of cross-cultural studies (Taylor, 1994).

Given the educational differences in South Africa, Taylor (1997) suggests the learning potential approach to identify those who have potential for development and who could benefit from development programmes. Proponents of this approach argue that cognitive ability is not fixed or static; rather, it can change. Following Vygotsky's (1978) views, this approach supports the idea that performance on its own is not a true reflection of a person's cognitive ability. One must assess the ability to learn in the context of mediation, if one wants a true reflection of people's abilities. Potential is assessed rather than only current abilities or skills and this orientation toward the future to some extent avoids a problem in ability testing which reflects an individual's past opportunities for and experiences with learning. If these have not been conducive to the development of the abilities measured by the ability test in question the individual will be at a disadvantage relative to others who had better educational opportunities. Test developers however still need to address possible bias and Claassen (1997) questions whether these tests discriminates fairly among blacks. South African learning potential assessment tools include the APIL, TRAM-1 and TRAM-2 developed by Taylor (1997). Care has been taken to limit cultural content in the stimulus material of these instruments and to ensure that the material is novel to everyone. The results of predictive bias studies have been positive and the instruments seem to predict educational and training criteria (Taylor, 1997). According to De Beer (1994), using a test-train-retest model as in another learning potential instrument, LPCAT, is a means of obtaining a more comprehensive picture of the person's current developmental state and his/her potential future level of ability. Shochet (1994) also demonstrated the possible utility of the learning potential approach in improving the predictive efficiency of admission procedures. In his study it was found that cognitive modifiability of applicants to a tertiary educational facility significantly moderated predictive efficiency of the traditional intelligence measure used in admissions selection.

THE MEANING SHIFT SUSCEPTIBILITY OF PERSONALITY TESTS

Retief (1987) suggests that the explicit structure of objective, factorially based methods of personality assessment may make them less suitable as cross-cultural instruments than projective tests, because the structure of these methods rigidly articulates the system of meaning of the test constructor and makes them more susceptible than projective methods to cultural shifts in meaning attributed to events and situations. He notes that it is apparent from the literature that factorially based personality tests rarely retain the level of reliability when applied cross-culturally (Retief, 1988). Furthermore, sometimes the validity diminishes substantially. This, however, does not imply that projective techniques are not also subject to cultural variation. Traditional personality tests rely on the method of depicting situations where different action options are available, or where different interpretations of events reveal aspects of personality. The 'meaning shift susceptibility' of personality tests also rests on the fact that the typical questions in them concern events in the domain of social interaction, a domain that is always, in any culture, significantly influenced by action prescriptions that flow from a culture's broader value system.

Taylor and Boeyens (1990) give a brief review of the cross-cultural research on personality measures. This research gives a somewhat mixed impression of the performance of questionnaires in cultures for which they were not designed. The South African studies have produced findings that led Taylor and

Boeyens to question whether the traits measured actually exist as viable constructs in black cultures. Based on their study with the South African Personality Questionnaire (*SAPQ*) they concluded that while there was some support for construct comparability between blacks and whites, analysis of item bias indicated that the questionnaire was not suitable for cross-cultural application. Although Retief (1992) agrees that the *SAPQ* should not be used in a multicultural context, he points out that bias in the context of personality tests can also be useful if it is viewed as difference, rather than as deficit. The latter view stems from the importation of bias detection techniques from the cognitive functioning domain, where bias can imply discrimination. Since personality test items deal with interpretation of social situations and events, subjects can be expected to differ in terms of the meanings they ascribe to such items, and such meanings are not necessarily quantifiable. Thus, Retief argues that it is perhaps misleading to equate differential item endorsement and incomparability (i.e. difference with deficit).

Holtzman (in Retief, 1988) points out that there are a number of problems with personality testing relating mainly to the difficulty in establishing reliability and validity. These result from the free-response nature of the task, and the high degree of interaction between situational factors and personality, as well as the difficulty in obtaining suitable independent criteria against which to validate inferences about unconscious aspects of personality. Further problems include cross-cultural transmission of meaning in personality assessment, situational and instrument factors, maintaining semantic equivalence across cultures, and delineation and control of relevant cultural variables.

According to Van Eeden, Taylor, and Du Toit (1996) the meaning of certain phrases and words in the Sixteen Personality Factor Questionnaire Fifth Edition (*16PF5*), may not be clear to individuals who do not have English as first language. Abrahams (1997) has similar concerns with regard to the Sixteen Personality Factor Questionnaire, South African 1992 version (*16PF, SA92*). She also recommends that this questionnaire should not be used cross-culturally as the reliability was not acceptable for all the groups included in her study. Although the results of Van Eeden and Prinsloo (1997) were generally more positive, they found some differences in the loading patterns on the second order factors of the *16PF, SA92* for blacks and whites. These findings may suggest useful information, but the results have implications for the use of the test in a multicultural context. They concluded that while separate norms for different population groups appear not to be necessary, certain cultural and gender specific trends must be accounted for when interpreting results. Along with Taylor and Boeyens (1990) the authors mentioned above recommend further study of variables such as language proficiency, level of education, socio-economic status, degree of urbanization, and values (among other things) as these have been shown to influence responses to personality test items. In the current standardization of the South African English version of the *16PF5* (Prinsloo, 1998), measurement equivalence as well as the effect of language proficiency on personality profiles are being studied. The research by Heuchert (1997) on the NEO Personality Inventory – Revised (*NEO-PI-R*) is another example of efforts to investigate the possible impact of culture on the reliability and validity of test results. This type of research aims to assess the applicability of the five factor model of personality to South Africa.

Fair practices in South Africa might require a greater emphasis on work-related variables (Abrahams, 1994). Given the difficulties experienced in cross-cultural testing, Abrahams (1994) recommends that only information which is systematically related to work success should be considered for decision making. Criteria include, but are not limited to the ability to perform tasks. Although the Occupational Personality Questionnaire (*OPQ*) is a psychometric test, it focuses on competencies that are of importance in occupational contexts. Competencies can be seen as aspects of a person, such as inter-

personal sensitivity, that enables him or her to be competent in an area that has been identified as important for specific posts. To determine if the *OPQ* predicts occupational behaviour, Robertson and Kinder (1993) did a meta-analysis testing specific hypotheses. This study provided strong evidence of the criterion-related validity of the *OPQ*. The inventory is currently used by a number of organizations in South Africa. Saville and Holdsworth South Africa is using this database to determine the validity of the *OPQ* in a multicultural context. Similarly the Psychological Assessment Initiative (PAI) aims to make use of available data from industry to check the validity of tests being used in specific contexts. This initiative was started at the 1997 conference of the Society for Industrial Psychology who in association with PAI has published guidelines for the validation and use of assessment procedures in the workplace.

FAIR USE OF TESTS

Psychological test results are used in educational settings, in clinical practice and for selection and placement in industry. It is important for these various purposes that tests demonstrate adequate validity. Owen (1996) notes in this regard that South African tests are generally reliable and valid, but that this applies mainly to the groups for which they were standardized. The test constructor should report on the psychometric properties of a test but it is the responsibility of the test user to study the relevant information and to determine if a test complies with the necessary requirements for the group the test is administered to. The test user may also be expected to examine the validity of tests for the various contexts in which it is being used (the PAI is an example of such an initiative). From the foregoing discussion, it should be apparent that establishing comparability of scores across groups is vital in a country where many different people from a variety of cultural groups compete for various opportunities. It is necessary in this context to distinguish between test bias (as discussed above) and fair use of tests and their results.

Test bias is determined by means of objective, statistical indices that indicate if the test scores have the same or different meanings and implications for different subgroups. Test fairness on the other hand reflects social values and philosophies of test use. Regardless of the statistical properties of the test, test results will be used differently depending on the ethical position taken by the selector and in each instance various factors including the social consequences will determine if the decision is regarded as fair or not.

The issue of fairness in testing can be related to a concern for securing equal opportunity for all. Thus, fairness can be viewed as acceptance of the candidate identified as most likely to be successful in a given position. It can on the other hand refer to the extent that selection procedures succeed in meeting certain socio-political goals, especially those pertaining to the minimization of inequality of opportunity and increasing the representativeness of disadvantaged demographic groups (Reynolds, in Huysamen, 1996). Taylor and Radford (1986) argue that any interpretation of psychometric data in South Africa that does not take into account possible differences between ethnic groups is likely to ignore a significant moderator variable, given the history of ethnically based discriminatory practice in this country.

Flowing from the concept of unfairness has been the creation of a number of 'fair selection models' which should be used when a test functions differently in different groups. These models include the regression model, the quota model, the equal risk model, the constant ratio model, and the conditional probability model. Although these models have been developed to assist fair selection, there is no rationale for choosing between them (Owen, 1996; Taylor & Radford, 1986). In tertiary educational selections the decision must also be mediated by the consideration of financial resources available to subsidize students. It is necessary in this setting to consider the cost

(in terms of both money and potential dropping of standards) of admitting applicants less likely to succeed at the expense of those predicted to do well (Huysamen, 1996). However, such placements are necessary if educational opportunity is to be equalized. Huysamen (1996) suggests that demographic information should be considered together with psychological tests when developing admission procedures. Essentially, the search for a fair selection procedure involves a trade-off between merit-only selection and selection that addresses the problem of representativeness (Taylor & Radford, 1986).

Another concern is how tests and testing are perceived by workers, personnel practitioners, and other stakeholders in the testing process. Vorster (1983) reports that problems identified in testing in industry include the acceptability of tests and testing, test format, relevance of tests to the job, interpretation of results, quality of instructions, adequacy of testers, and the purpose of the testing. Similarly, Shaw and Human (1989), and Sehlapelo and Terre Blanche (1996) suggest that generally, workers included in their studies were highly critical of many tests and testing practices. Problems identified included the following: Constructs measured were considered irrelevant to the job being applied for, testing was perceived as dehumanizing, and there was felt to be a lack of feedback to applicants. Workers tended to have little trust in tests and the testing process, the general consensus (from practitioners and workers) being that tests are unfair, irrelevant and/or biased against blacks.

Shaw and Human (1989) note that clear criteria for assessment are required, and the assessment of potential must proceed in relation to a clearly defined position. In addition, assessment should incorporate the recognition of dynamic change in the potential of candidates. Sehlapelo and Terre Blanche (1996) indicate that workers saw the solution to the problems in testing in the formulation of explicit testing policies arrived at through a process of consultation. On the other hand, personnel practitioners tended to think that the cultural bias in testing could be overcome through construction of non-biased tests.

In a similar vein, Magwaza (1995, p.12) notes that "culture-fairness has become a rhetoric too nebulous to define". It is argued that this obsession with test bias precludes a focus on making the testing context more human and democratic. To this end there is a need for testers (especially in group contexts) to become culturally sensitive, and to deal with issues like the power dynamics of the testing situation, and possible negative 'baggage' that testees may bring to the testing context (based on previous experiences with testing). Furthermore, Magwaza points to the need for better feedback, to give testees the opportunity to address weaknesses identified by the testing. The author concludes that "outcomes of any assessment . . . will only be meaningful . . . if a facilitative environment is created for participants to unleash their manifest and latent potential" (Magwaza, 1995, p.15).

As English has become the lingua franca of business and government, a greater responsibility lies with employers to ensure an adequate proficiency in this language among their employees (Price, 1997). A reasonable level of English proficiency is required to avoid conflict and misunderstanding, and promote productivity. As has been mentioned previously, this improvement in language skills is likely to prove beneficial to the test performance of many African language speakers.

CONCLUSION

There is a clear awareness among those working in the field of psychometric testing of a continued need for change in order to make those tests that are available, as well as those to be developed in the future, relevant for the multicultural society that is South Africa. To this end a number of studies have been done focusing on the validity of various tests used in the educational, clinical and industrial spheres. In general, these

studies support the view that South African tests are reliable and valid for the groups for which they were developed and standardized. This means that comparison of scores of individuals within various groups may be justified. However, the fact that cross-cultural validity has not been unequivocally determined for a number of tests implies that cross-group comparison of scores could yield information that is discriminatory if used unconditionally.

Claassen (1997, p. 306) states that: "never can a test score be interpreted without taking note of and understanding the context in which the score was obtained". Test users should consider the role of factors such as socio-economic status, educational level and quality, language proficiency, and acculturation (among others) in moderating test performance of various groupings of people in South Africa. This implies utilizing existing research findings as well as continued research in this regard. For example, in the study by Van Eeden and Prinsloo (1997) certain trends possibly related to cultural differences were observed in the personality profiles of the different groups. When testing in that specific context, individual results can be interpreted taking these trends into account. In another example of relying on experience and research findings, Shuttleworth-Jordan (1996) illustrates the use of a clinical analysis of results (including psychometric test results) as an alternative to a normative-based approach in cross-cultural assessment.

As the needs and agenda of those associated with testing have changed with time so the recognition of culture and the variables it encompasses has changed. In South Africa this has been exemplified by recognition of the role of past discriminatory practices in current differential performance of various groups on psychometric tests. Concomitantly there has been a move towards addressing past wrongs. It is a misconception that psychometric test results represent an 'inherent' and unchangeable ability or that these results are a hundred percent accurate and perfectly reliable (Owen, 1998). Tests cannot supply all the answers, no matter how justifiable the construction process. As test users have been forced to evaluate the usefulness and necessity of testing, a shift towards multi-method assessment has occurred (Foxcroft, 1997b). As such, practitioners increasingly acknowledge that test results are only a part of the information that can be obtained for the purposes of intervention, selection, guidance, or placement. Visser (1996) concludes that decisionmaking in these areas also involves ideological and social choices.

ACKNOWLEDGEMENT

This literature review was carried out by the first author as part of an internship programme for the directed Master's degree in Research Psychology at the University of South Africa. The authors wish to thank two anonymous reviewers for their helpful comments on an earlier draft of this manuscript.

REFERENCES

- Abrahams, F. (1994). *The cross-cultural comparability of personality tests in South Africa: Is it possible?* Paper presented at the Congress of Psychometrics, Johannesburg.
- Abrahams, F. (1997). *Problems associated with the continued use of existing psychological tests in a new South Africa. A study using the Sixteen Personality Factor Questionnaire (16PF) as an example.* Paper presented at the 3rd Annual Congress of the Psychological Association of South Africa, Durban.
- Baron, R.M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Claassen, N.C.W. (1990). The comparability of General Scholastic Aptitude Test scores across different population groups. *South African Journal of Psychology*, 20, 80-91.
- Claassen, N.C.W. (1996). *Paper and Pencil Games (PPG): Manual.* Pretoria: Human Sciences Research Council.

- Claassen, N.C.W. (1997). Cultural differences, politics and test bias in South Africa. *European Review of Applied Psychology*, 47(4), 297-307.
- Claassen, N.C.W., & Cudeck, R. (1985). Die faktorstruktuur van die Nuwe Suid-Afrikaanse Groeptoets (NSAG) vir verskillende bevolkingsgroepe (The factor structure of the New South African Group Test (NSAGT) for different population groups). *South African Journal of Psychology*, 15, 1-10.
- Claassen, N.C.W., & Schepers, J.M. (1990). Groepverskille in akademiese intelligensie verklaar op grond van verskille in sosioekonomiese status (Group differences in academic intelligence explained by virtue of differences in socio-economic status). *South African Journal of Psychology*, 20, 294-302.
- De Beer, M. (1994). *Dynamic assessment – meeting new challenges in training and development*. Paper presented at the 6th conference of the Southern African Institute for Management Scientists, Cape Town.
- Ferguson, G. (1956). On transfer and the abilities of man. *Canadian Journal of Psychology*, 10, 121-131.
- Foxcroft, C. (1997a). *The PASS theory of cognitive processes as an alternative model in intellectual assessment*. Paper presented at the National Symposium on the standardisation of the WAIS III for South Africa, Pretoria.
- Foxcroft, C. (1997b). Psychological testing in South Africa: Perspectives regarding ethical and fair practices. *European Journal of Psychological Assessment*, 13, 229-235.
- Heuchert, J.W.P. (1997). *The applicability of the five factor model of personality in a South African sample*. Paper presented at the 3rd Annual Congress of the Psychological Association of South Africa, Durban.
- Holburn, P.T. (1992). *Test bias in the Intermediate Mental Alertness, Mechanical Comprehension, Blox, and High Level Figure Classification tests* (Contract Report C/Pers-453). Pretoria: Human Sciences Research Council.
- Hugo, H.L.E., & Claassen, N.C.W. (1991). *The functioning of the GSAT Senior for students of the Department of Education and Training* (Report ED-13). Pretoria: Human Sciences Research Council.
- Huysamen, G.K. (1996). Fair and unbiased admission procedures for South African institutions of higher education. *South African Journal of Higher Education*, 10(2), 199-207.
- Irvine, S.H., & Berry, J.W. (1988). The abilities of mankind: A re-evaluation. In S.H. Irvine & J.W. Berry (Eds.), *Human abilities in cultural context* (pp. 3-59). Cambridge: Cambridge University Press.
- Kendall, I.M., Verster, M.A., & Von Mollendorf, J.W. (1988). Test performance of blacks in South Africa. In S.H. Irvine & J.W. Berry (Eds.), *Human abilities in cultural context* (pp. 299-339). Cambridge: Cambridge University Press.
- Magwaza, S. (1995). Humanising and democratising assessments. *People Dynamics*, 13(5), 12-15.
- Nell, V. (1994). Interpretation and misinterpretation of the South African Wechsler-Bellevue Adult Intelligence Scale: A history and a prospectus. *South African Journal of Psychology*, 24, 100-109.
- Nell, V. (1997). Science and politics meet at last: The insurance industry and neuropsychological norms. *South African Journal of Psychology*, 27, 43-49.
- Nell, V. (1999). *Cross-cultural neuropsychological assessment: Theory and practice*. Lawrence Erlbaum (in press).
- Owen, K. (1989). *Test and item bias: The suitability of the Junior Aptitude Tests as a common test battery for white, Indian and black pupils in Standard 7* (Report P-96). Pretoria: Human Sciences Research Council.
- Owen, K. (1991). Test bias: The validity of the Junior Aptitude Tests (JAT) for various population groups in South Africa regarding the constructs measured. *South African Journal of Psychology*, 21, 112-118.
- Owen, K. (1992). The suitability of Raven's Standard Progressive Matrices for various groups in South Africa. *Personality and Individual Difference*, 13(2), 149-159.
- Owen, K. (1996). Test bias and test fairness. In K. Owen & J.J. Taljaard (Eds.), *Handbook for the use of psychological and scholastic tests of the HSRC* (pp. 77-96). Pretoria: Human Sciences Research Council.
- Owen, K. (1998). *The role of psychological tests in education in South Africa: Issues, controversies and benefits*. Pretoria: Human Sciences Research Council.
- Poortinga, Y.H., & Van der Flier, H. (1988). The meaning of item bias in ability tests. In S.H. Irvine & J.W. Berry (Eds.), *Human abilities in cultural context* (pp. 166-183). Cambridge: Cambridge University Press.
- Price, H. (1997). The value of managing language diversity. *People Dynamics*, 15(1), 20-22.
- Prinsloo, C.H. (1998). *The factor structure and readability of the S.A. English version of the Fifth Edition of the American 16PF* (Unpublished report). Pretoria: Human Sciences Research Council.
- Retief, A.I. (1987). Thematic apperception testing across cultures: Tests of selection versus tests of inclusion. *South African Journal of Psychology*, 17, 47-55.
- Retief, A.I. (1988). *Method and theory in cross-cultural psychological assessment*. Pretoria: Human Sciences Research Council.
- Retief, A.I. (1992). The cross-cultural utility of the SAPQ – bias or fruitful differences? *South African Journal of Psychology*, 22, 202-207.
- Robertson, I.T., & Kinder, A. (1993). Personality and job competencies: The criterion-related validity of some personality variables. *Journal of Occupational and Organizational Psychology*, 66(3), 225-244.
- Sehlapelo, M., & Terre Blanche, M. (1996). Psychometric testing in South Africa: Views from above and below. *Psychology in Society*, 21, 49-59.
- Shaw, D., & Human, L. (1989). Assessment of managerial potential among black employees in selected South African companies: Methods, perceived problems and concerns. *South African Journal of Labour Relations*, 13(2), 4-19.
- Shochet, I.M. (1994). The moderator effect of cognitive modifiability on traditional undergraduate admissions test for disadvantaged black students in South Africa. *South African Journal of Psychology*, 24, 208-215.
- Shuttleworth-Jordan, A.B. (1995). Against mutton masquerading as lamb: Additional commentary on the interpretation and misinterpretation of the SAWAIS. *South African Journal of Psychology*, 25, 244-246.
- Shuttleworth-Jordan, A.B. (1996). On not reinventing the wheel: A clinical perspective on culturally relevant test usage in South Africa. *South African Journal of Psychology*, 26, 96-102.
- Taylor, T.R. (1994). A review of three approaches to cognitive assessment, and a proposed integrated approach based on a unifying theoretical framework. *South African Journal of Psychology*, 24, 184-191.
- Taylor, T.R. (1997). *Learning potential: An alternative to conventional assessment*. Paper presented at the National Symposium on the standardisation of the WAIS III for South Africa, Pretoria.
- Taylor, T.R., & Boeyens, J. (1990). *A comparison of black and white responses to the South African Personality Questionnaire* (NIPR Report PERS-440). Pretoria: Human Sciences Research Council.
- Taylor, J.M., & Radford, E.J. (1986). Psychometric testing as an unfair labour practice. *South African Journal of Psychology*, 16, 79-86.
- Tollman, S.G., & Msengana, N.B. (1990). Neuropsychological assessment: Problems in evaluating the higher mental functioning of Zulu-speaking people using traditional western techniques. *South African Journal of Psychology*, 20, 20-24.
- Van den Berg, A.R. (1985). *Using the Junior South African Individual Scales (1981) for testees from South African population groups which were not included in the norm population*. Pretoria: Human Sciences Research Council.
- Van den Berg, A.R. (1996). Intelligence tests. In K. Owen & J.J. Taljaard (Eds.), *Handbook for the use of psychological and scholastic tests of the HSRC* (pp. 157-190). Pretoria: Human Sciences Research Council.
- Van Eeden, R. (1993). *The validity of the Senior South African Individual Scale – Revised (SSAIS-R) for children whose mother tongue is an African language: Private schools*. Pretoria: Human Sciences Research Council.

- Van Eeden, R., & Prinsloo, C.H. (1997). Using the South African version of the 16PF in a multicultural context. *South African Journal of Psychology*, 27, 151-159.
- Van Eeden, R., Taylor, T.R., & Du Toit, R. (1996). *Adaptation and standardization of the Sixteen Personality Factor Questionnaire Fifth Edition (16PF5) in South Africa: A feasibility study* (Unpublished report). Pretoria: Human Sciences Research Council.
- Van Eeden, R., & Van Tonder, M. (1995). *The validity of the Senior South African Individual Scale - Revised (SSAIS-R) for children whose mother tongue is an African language: Model C schools*. Pretoria: Human Sciences Research Council.
- Van Eeden, R., & Visser, D. (1992). The validity of the Senior South African Individual Scale -Revised (SSAIS-R) for different population groups. *South African Journal of Psychology*, 22, 163-171.
- Verster, J.M., & Prinsloo, R.J. (1988). The diminishing test performance gap between English speakers and Afrikaans speakers in South Africa. In S. H. Irvine & J. W. Berry (Eds.), *Human abilities in cultural context* (pp. 534-559). Cambridge: Cambridge University Press.
- Viljoen, G., Levett, A., Tredoux, C., & Anderson, S. (1994). Using the Bender Gestalt in South Africa: Some normative data for Zulu children. *South African Journal of Psychology*, 24, 145-151.
- Visser, D.J. (1996). *Navorsingsgeleenthede vir gedragswetenskaplikes in die werkplek* (Research opportunities for behavioural scientists in the work place). Professorial inaugural address, Rand Afrikaans University.
- Vorster, J.F. (Ed.), (1983). *Simposium oor die problematiek wat ontstaan by die gebruik van dieselfde of afsonderlike toetse vir verskillende bevolkingsgroepe* (Symposium on the problems resulting from the use of the same or separate tests for different population groups) (Geleendheidspublikasie No. 6). Pretoria: Human Sciences Research Council.
- Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.