

Investigating subscale differences among race and language groups on the Occupational Personality Profile

Lindsey Joseph

University of Cape Town, South Africa

Burger van Lill

Section of Organisational Psychology, School of Management Studies, University of Cape Town, Private Bag, Rondebosch, 7701, South Africa
burger.vanlill@uct.ac.za

Our purpose in this study was to determine whether results of the subscales of the Occupational Personality Profile (OPP) are comparable among various race and language groups. The sample consisted of 234 individuals, who had applied for management positions in various government institutions. The sample was divided into African and white individuals from different language backgrounds. Analyses revealed acceptable reliability coefficients on most of the subscales of the OPP. Significant differences among means were however found on a number of the subscales. The practical implications of these findings for employment testing and adverse impact are discussed.

Keywords: adverse impact; employment testing; Occupational Personality Profile; personality testing; psychological testing; subscale differences

In today's ever changing business environment, organisations need to invest in employing and developing people with the necessary skills and competencies to become and remain competitive. It is therefore essential to identify behaviours and abilities that can be used to predict job success (La Grange & Roodt, 2001). As employers acknowledge the need for competent and skilled employees, they are also realising the value of psychological testing as an aid in obtaining and retaining the best people (Johnson & Kleiner, 2000).

Extensive research has clearly demonstrated that psychological testing can aid in predicting job performance (Barrick, 2001; Barrick & Mount, 1991; Barrick, Mount & Judge, 2001; Chamorro Premuzic & Furnham, 2003; La Grange & Roodt, 2001; Lievens, Coetsier, De Fruyt & De Maeseneer, 2002; Salgado, 1997; Schmidt, Ones & Hunter, 1992; Van der Walt, Meiring, Rothmann & Barrick, 2002). By providing valid and reliable evidence of cognitive functioning, intellectual abilities and personality characteristics, psychological testing is a valuable tool for selection and placement. However, whilst the value of psychological testing is well established, there is a critical need for empirical research on the extent to which imported psychological tests can be used to make comparisons between various demographic groups within our local context. Many adapted versions of imported tests have been proved to be problematic when used in South Africa. Because employers need to consider ethical and legal implications associated with the use of any particular measuring instrument, it is important to assess group differences such as race and language on the psychological tests commonly used in industry (Packman, Brown, Englert, Sisarich & Bauer, 2005). Very often selection and promotion decisions are made on the basis of tests that have not been proved to be comparable across different racial and language groups (Abrahams, 1996).

Developments in South African labour legislation, and in particular the Employment Equity Act 55 of 1998 (EEA), compel test validation, specifically in industry. However, before the promulgation of the EEA, black people (Africans, coloureds and Indians) were not protected from unfair discrimination, which was often exacerbated by the misuse and abuse of tests and test results (Abrahams, 1996). Psychological testing in South Africa should therefore not be considered without reflecting

on the past discriminatory laws and practices of apartheid. In fact, large inequalities in South Africa's social and economic structure still exist today. As a result, variables such as language, race, social and educational backgrounds are all likely to have an influence on an individual's test performance.

The main purpose in the present research was therefore to investigate whether the results of the Occupational Personality Profile (OPP), a personality test commonly used in South Africa for personnel selection and placement, are comparable among various racial and language groups. According to Ones and Anderson (2002), analysing group differences associated with personality tests can provide insight into whether a particular group is being adversely impacted by the use of such tests.

Psychological testing in South Africa

From about 1920 to 1960, psychological tests used were either adaptations of foreign tests or were developed specifically for the white population. Test results were often used to draw conclusions about differences among racial groups without investigating the appropriateness of the instrument. From the 1960s however, there was a growing recognition that race and language are important moderators of test performance, and consequently the assessment of black South Africans became more systematic and thorough (Bedell, Van Eeden & Van Staden, 1999). Major developments in the manufacturing and mining industries led to the placement of large numbers of black semi skilled workers, which required a more rigorous, unbiased approach. Following this period, the development of tests of adaptability dominated interest (Bedell *et al.*, 1999). Certain aspects of fairness, bias and discriminatory practices received more attention in line with international developments (Meiring, Van de Vijver, Rothmann, & Barrick, 2005). Greater awareness of cultural bias inherent in the tests themselves furthermore strengthened the notion that culture may constitute a source of systematic error in test results (Bedell *et al.*, 1999). From about 1980 to 1994, there was a growing demand for tests that were more appropriate for different race and language populations.

The first thorough study of bias nonetheless only took place when Owen (1986) investigated test and item bias using the Senior Aptitude Test, the Mechanical Insight Test and the Scholastic Proficiency Test. He found significant differences between the test scores of black and white individuals and concluded that understanding the reasons for such differences and reducing them would be a major challenge. Taylor and Boeyens (1991) conducted their study on the South African Personality Questionnaire and concluded that although there was some evidence of comparability between black and white individuals, some items were so biased that the cross cultural application of the test would be problematic. Retief (1992) also noted that personality tests seldom retain the level of reliability when used across cultures and the validity sometimes even diminishes substantially.

Abrahams (1996) and Abrahams and Mauer (1999) conducted studies on the Sixteen Personality Factor Questionnaire (16PF) and they too concluded that the 16PF could not be used across different racial groups, as the reliability was not acceptable for the black groups. The results suggested problems with the construct and item comparability of the 16PF and significant mean differences were found when the different race groups were compared. Abrahams (2002) asserted that thorough research must be conducted on any imported psychological test, prior to its adoption within South Africa.

Considering the history of South Africa's language policies and differences in language proficiencies, administering a psychological test in English can become problematic for individuals from a particular demographic group. The lack of English proficiency may in fact have a negative influence on an individual's performance on a test (Meiring, Van de Vijver & Rothmann, 2006). It is therefore important to consider language when evaluating the appropriateness of a psychological test in a multi lingual context (Van de Vijver & Leung, 1997). Abrahams (2002) established that participants whose home language was neither English nor Afrikaans found it more difficult to understand the items of the 16PF. Meiring (2000) also found significant differences between racial groups on the

16PF and attributed this to the fact that the black individuals did not adequately understand the items. Meiring *et al.* (2005) conducted a study on the construct, item, and method bias of the Fifteen Factor Questionnaire Plus (15FQ+) in South Africa. Their study revealed construct bias for various demographic groups. Low internal consistencies for several personality scales were also revealed, specifically in the black groups. In an investigation into bias in an adapted version of the 15FQ+ it was further concluded that although there were improvements in the adapted version, there were almost no significant improvements in internal consistencies for the black groups (Meiring *et al.*, 2006). In a subsequent study examining bias in the 15FQ+ and the OPP, Meiring (2007) found low internal consistencies for both tests.

The above findings provide useful information and have significant implications for the use of personality tests in South Africa. The EEA imposes very strict requirements for the use of any psychological test or any similar form of assessment. It seems that race and language may moderate test scores to such an extent that they affect the probability of an individual being selected for employment. In a study conducted by Foxcroft, Paterson, Le Roux and Herbst (2004) on test utilisation patterns and the needs of psychological assessment practitioners in South Africa, practitioners emphasised the need to adapt tests for South Africa's diverse society and to address language issues in testing in particular. Despite the need for culturally relevant tests, many personality tests used in South Africa today have been imported from the UK and USA. The lack of locally developed tests can be attributed in part to a significant shortage of experienced test developers in South Africa (Foxcroft, 2004). Therefore, if imported tests are to be used, it becomes important to establish whether these tests are in fact suitable for a diverse and multicultural society such as South Africa. Examining the requirements as set out in the EEA, and considering the studies referred to above as well as the extent to which personality testing is used in industry today, there are concerns regarding whether the requirements of the EEA for some personality tests are in fact being met.

Therefore, we examine the differences among the OPP scale means of various demographic groups who were assessed for selection purposes. Consistent with previous research, the study is not an investigation of test bias, but aims to ascertain the possibility of adverse impact by examining trait differences across groups (*cf.* Goldberg, Sweeney, Merenda & Hughes, 1998; Ones & Anderson, 2002; Ones & Viswesvaran, 1998; Packman *et al.*, 2005).

RESEARCH DESIGN

Sample

The sample comprised 234 individuals who had applied for middle to senior management positions across various government institutions. The respondents were from a range of occupations and industries and were all in possession of a postgraduate qualification.

The sample included African ($n = 169$) and white ($n = 65$) respondents, and their mean age was 40.75, ranging from 22 to 67 years. Sixty six percent ($n = 155$) were male and 34% ($n = 79$) were female. The sample was further divided into an English language group ($n = 39$) and an Afrikaans language group ($n = 20$). Additional language groups included Sepedi, Sesotho, Setswana, siSwati, Tshivenda, Xitsonga, Ndebele, Xhosa, and Zulu. However, for the purposes of analysis, these smaller language groups were combined to form the African language group ($n = 99$). The 32% ($n = 76$) of respondents who had not indicated their home language on the application forms were treated as missing data.

Measuring instrument

The OPP is a personality test designed for personnel assessment and selection purposes. The test was originally developed in the UK and was introduced in South Africa in 1995 (Tredoux, n.d.). The OPP has since been classified as a psychological test by the Psychometrics Committee of the Health

Professions Council of South Africa. It is recommended that individuals who complete the OPP must be in possession of a Grade 12 level of education and be proficient in English. The OPP comprises 98 test items using a five point response scale (strongly disagree to strongly agree). The test measures nine primary content scales in addition to a motivational distortion scale. The nine scales that form the OPP are detail conscious flexible, genuine persuasive, reserved gregarious, cynical trusting, optimistic pessimistic, emotional phlegmatic, accommodating assertive, composed contesting, and abstract pragmatic. Detailed descriptions are provided in Table 1.

Table 1. Descriptions of the OPP scales

Detail Conscious Meticulous, rule bound, rigid, conscientious, detail orientated, conventional	Flexible Unconventional, flexible, dislikes detail, lacks self discipline, disregards rules and obligations
Genuine Forthright, honest and open, genuine and sincere, may lack tact and diplomacy	Persuasive Diplomatic, persuasive, socially aware, may be shrewd and calculating
Reserved Introspective, aloof, prefers to work alone, enjoys own company	Gregarious Sociable, high energy, enjoy working with others, high need for affiliation
Cynical Suspicious, sceptical, cautious, may distrust others, inclined to question others motives	Trusting Takes people at face value, have faith in other's honesty, sometimes credulous and naïve
Optimistic Achievement orientated, have faith in their own ability, positive approach to set backs, believe they are in control of their own destiny	Pessimistic Resigned, have little faith in their ability to determine events, prone to feelings of helplessness, fatalistic, negative approach to setbacks
Emotional Anxious, temperamental, prone to suffer from feelings of self doubt and may have difficulty coping under pressure	Phlegmatic Self assured, emotionally stable and secure, resilient, capable of dealing effectively with demanding situations
Accommodating Accepting, sensitive to other's feelings, avoids confrontation, modest and unassuming	Assertive Dominant, task orientated, challenging, confrontative, unconcerned about others feelings
Composed Calm, tolerant, prefer working at a steady pace, dislike having to continually meet close deadlines, do not have an aggressive, competitive attitude	Contesting Ambitious and competitive, work long hours, rush to meet deadlines, impatient, may take on too much work, prone to stress related illnesses
Abstract Imaginative, creative, aesthetically sensitive, have a theoretical orientation, little regard for practical matters	Pragmatic Concrete, practical and realistic, not interested in artistic matters, more concerned with 'how than why'

Note: The nine scales measured by the OPP are bi polar. Descriptions in the left column of the table indicate low scores, whilst descriptions on the right of the table are indicative of high scores. (Psytech International, n.d, pp.20 28)

The personality scales measured by the OPP were included as a result of their relevance to occupational assessment and decision making and because of the extensive research literature supporting their reliability and validity (Psytech International, n.d). The contributions of several well known psychologists and researchers in the area of personality testing, such as Guilford, Cattell, Kline, Eysenck, Thurstone, Maslow, and Jung, were used in constructing the nine scales of the OPP. The Technical Manual (Psytech International, n.d) provides further details on the theoretical basis for the OPP dimensions.

Reliability and validity indices and information are reported in the Technical Manual (Psytech International, n.d.) and the South African User's Manual (Tredoux, n.d.). With the User's Manual, approximately 40 reliability studies were conducted on various groups, with the majority of the reliability studies coming from the South African Police Services. Table 2 indicates the reliability coefficients of three groups most similar to the present sample, where Alpha coefficients range from .49 to .79. The Alpha coefficients calculated for the present sample are presented in the Results section and in Table 5.

The User's Manual provides additional information on the reliability for applicants pre screened for English proficiency and applicants not pre screened for English proficiency for the South African Police Services group. In general, applicants who were not pre screened for English proficiency consistently obtained lower reliability scores. One can therefore deduce that English proficiency does play a role on an individual's performance on the OPP. It should however be noted that pre screening for English was not done in the present study.

Table 2. Internal consistency indices (Coefficient Alpha) for a variety of groups

OPP Subscale	Consultants/Candidates of all races for professional positions	SA applicants to postgraduate business school	Applicants for MBA courses
Detail Conscious Flexible	.67	.73	.65
Genuine Persuasive	.67	.72	.68
Reserved Gregarious	.55	.63	.63
Cynical Trusting	.70	.73	.77
Optimistic Pessimistic	.71	.63	.67
Emotional Phlegmatic	.66	.67	.79
Accommodating Assertive	.62	.69	.70
Composed Contesting	.73	.73	.60
Abstract Pragmatic	.54	.61	.64
Motivational Distortion	.50	.56	.49

Procedure

All respondents in the study were required to complete the OPP as part of an assessment centre battery. They completed the assessment in groups of four or five between 2003 and 2005 for selection purposes. The tests were completed in pencil and paper form and were administered under standardised conditions under the guidance of trained psychometrists and psychologists. The required permission to conduct the research was obtained from the relevant organisations. Anonymity and confidentiality of individual scores were assured and the researchers did not have access to personal information.

RESULTS

Differences between language groups

To test for significant differences between the African, English and Afrikaans language groups,

one way ANOVA and Scheffé post hoc comparisons were used. Table 3 shows the table of descriptive statistics from the one way procedure. In cases where significant differences were found, effect sizes were calculated in terms of Cohen’s *d* values and their confidence limits (see Rosenthal & Rosnow (2008) for a discussion and relevant formulae). According to Rosenthal and Rosnow (2008, pp. 364 365) effect size can be interpreted as follows: *d* .2 is “small”, *d* .5 is “medium” and *d* .8 is “large”.

Table 3. Descriptive statistics and differences among OPP subscale means for language groups

OPP Subscale	Language group	<i>n</i>	Mean	<i>SD</i>	<i>F</i>	<i>df</i>	<i>p</i>
Detail Conscious Flexible	African	99	38.38	5.06	6.53	2, 176	.002
	Afrikaans	20	35.55	4.51			
	English	60	35.73	4.96			
	Total	179	37.18	5.12			
Genuine Persuasive	African	99	28.22	4.35	5.56	2, 176	.005
	Afrikaans	20	30.35	2.68			
	English	60	26.80	4.51			
	Total	179	27.98	4.37			
Reserved Gregarious	African	99	28.27	5.65	.24	2, 176	.789
	Afrikaans	20	28.45	4.84			
	English	60	28.92	6.13			
	Total	179	28.51	5.71			
Cynical Trusting	African	99	29.54	5.91	8.89	2, 176	.000
	Afrikaans	20	27.00	5.59			
	English	60	25.35	6.66			
	Total	179	27.85	6.41			
Optimistic Pessimistic	African	99	29.00	5.06	7.53	2, 176	.001
	Afrikaans	20	30.90	4.10			
	English	60	31.93	4.22			
	Total	179	30.20	4.87			
Emotional Phlegmatic	African	99	38.86	5.57	2.15	2, 176	.119
	Afrikaans	20	40.55	5.24			
	English	60	40.68	6.19			
	Total	179	39.66	5.79			
Accommodating Assertive	African	99	27.14	4.51	3.37	2, 176	.037
	Afrikaans	20	30.30	5.51			
	English	60	27.40	5.59			
	Total	179	27.58	5.08			
Composed Contesting	African	99	31.60	6.59	3.43	2, 176	.035
	Afrikaans	20	27.85	5.84			
	English	60	30.20	5.36			
	Total	179	30.71	6.21			
Abstract Pragmatic	African	99	32.57	4.31	5.51	2, 176	.005
	Afrikaans	20	28.95	4.14			
	English	60	32.82	5.52			
	Total	179	32.25	4.85			
Motivational Distortion	African	99	23.11	3.80	.86	2, 176	.425
	Afrikaans	20	23.55	3.58			
	English	60	23.97	4.45			
	Total	179	23.45	4.00			

For the Detail Conscious Flexible scale, significant differences among the language groups were found, with $F(2, 176) = 6.53, p = .002$. In the post hoc comparisons, a significant difference was found between the mean of the African language group ($M = 38.38$) and the mean of the English language group ($M = 35.73$); $d = .53$ ($SD = .17, df = 157$) and the 95% confidence limits were .20-.86.

For the Genuine Persuasive scale, significant differences among the language groups were found, with $F(2, 176) = 5.56, p = .005$. In the post hoc comparisons, a significant difference was found between the mean of the English language group ($M = 26.80$) and the mean of the Afrikaans language group ($M = 30.35$); $d = .96$ ($SD = .27, df = 78$) and the 95% confidence limits were .41-1.50.

For the Reserved Gregarious scale, there were no significant differences among the language groups.

For the Cynical Trusting scale, significant differences among the language groups were found, with $F(2, 176) = 8.89, p < .001$. In the post hoc comparisons, a significant difference was found between the mean of the African language group ($M = 29.54$) and the mean of the English language group ($M = 25.35$); $d = .67$ ($SD = .17, df = 157$) and the 95% confidence limits were .33-1.00.

For the Optimistic Pessimistic scale, significant differences among the language groups were found, with $F(2, 176) = 7.53, p = .001$. In the post hoc comparisons, a significant difference was found between the mean of the African language group ($M = 29.00$) and the mean of the English language group ($M = 31.93$); $d = .63$ ($SD = .17, df = 157$) and the 95% confidence limits were .30-.96.

For the Emotional Phlegmatic scale, there were no significant differences among the language groups.

For the Accommodating Assertive scale, significant differences among the language groups were found, with $F(2, 176) = 3.37, p = .037$. In the post hoc comparisons, a significant difference was found between the mean of the African language group ($M = 27.14$) and the mean of the Afrikaans language group ($M = 30.30$); $d = .63$ ($SD = .25, df = 117$) and the 95% confidence limits were .13-1.12.

For the Composed Contesting scale, significant differences among the language groups were found, with $F(2, 176) = 3.43, p = .035$. In the post hoc comparisons, a significant difference was found between the mean of the African language group ($M = 31.60$) and the mean of the Afrikaans language group ($M = 27.85$); $d = .60$ ($SD = .25, df = 117$) and the 95% confidence limits were .11-1.10.

For the Abstract Pragmatic scale, significant differences among the language groups were found, with $F(2, 176) = 5.51, p = .005$. In the post hoc comparisons, a significant difference was found between the means of the African ($M = 32.57$) and Afrikaans ($M = 28.95$), $d = .86$ ($SD = .25, df = 117$) and the 95% confidence limits were .35-1.36; and the Afrikaans and English language groups ($M = 32.82$); $d = .79$ ($SD = .27, df = 78$) and the 95% confidence limits were .26-1.33.

For the Motivational Distortion scale, there were no significant differences among the language groups.

Differences between race groups

To provide a comparison, between the African and white racial groupings on the different scales of the OPP, t tests were used. The descriptive statistics are presented in Table 4.

For the Detail Conscious Flexible scale, significant differences were found between the mean of the African respondents ($M = 38.66$) and the mean of the white respondents ($M = 35.33$), with $t(232) = 4.71, p < .001$; $d = .68$ ($SD = .15, df = 232$) and the 95% confidence limits were .39-.98.

For the Genuine Persuasive scale, there were no significant differences between the race groups.

For the Reserved Gregarious scale, significant differences were found between the mean of the African respondents ($M = 26.96$) and the mean of the white respondents ($M = 29.18$), with $t(232) = 2.69, p = .008$; $d = .40$ ($SD = .15, df = 232$) and the 95% confidence limits were .11-.69.

For the Cynical Trusting scale, significant differences were found between the mean of the African respondents ($M = 28.52$) and the mean of the white respondents ($M = 26.12$), with $t(116) = 2.52, p = .013$; $d = .37$ ($SD = .15, df = 232$) and the 95% confidence limits were .08-.66.

For the Optimistic Pessimistic scale, significant differences were found between the mean of the African respondents ($M = 29.55$) and the mean of the white respondents ($M = 31.88$), with $t(148) = 3.70, p < .001$; $d = .51$ ($SD = .15, df = 232$) and the 95% confidence limits were .22-.80.

For the Emotional Phlegmatic scale, there were no significant differences between the race groups.

For the Accommodating Assertive scale, there were no significant differences between the race groups.

For the Composed Contesting scale, significant differences were found between the mean of the African respondents ($M = 32.56$) and the mean of the white respondents ($M = 28.63$), with $t(232) = 4.43, p < .001$; $d = .67$ ($SD = .15, df = 232$) and the 95% confidence limits were .37-.96.

Table 4. Descriptive statistics and differences between OPP subscale means for African and white respondents

OPP Subscale	Race group	<i>n</i>	Mean	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>
Detail Conscious Flexible	African	169	38.67	4.78	4.71	232	.000
	White	65	35.33	5.00			
Genuine Persuasive	African	169	27.19	4.81	.213	232	.831
	White	65	27.34	4.76			
Reserved Gregarious	African	169	26.96	5.74	2.69	232	.008
	White	65	29.18	5.44			
Cynical Trusting	African	169	28.52	6.50	2.52	116	.013
	White	65	26.12	6.53			
Optimistic Pessimistic	African	169	29.55	5.09	3.70	148	.000
	White	65	31.88	3.97			
Emotional Phlegmatic	African	169	40.10	5.79	1.53	232	.178
	White	65	41.26	6.12			
Accommodating Assertive	African	169	28.88	4.70	1.64	232	.103
	White	65	28.06	5.61			
Composed Contesting	African	169	32.56	6.28	4.43	232	.000
	White	65	28.63	5.48			
Abstract Pragmatic	African	169	32.49	4.29	.73	232	.469
	White	65	31.98	5.71			
Motivational Distortion	African	169	22.78	4.23	2.40	116	.018
	White	65	24.26	4.22			

Note: Appropriate *t* values and *df* selected in accordance with Levene’s test for equality of variance.

Table 5. Equality of reliability coefficients of the OPP subscales for African and white respondents

OPP Subscale	Race group	<i>n</i>	Alpha	<i>F</i>	<i>df</i>
Detail Conscious Flexible	African	169	.63	1.15	168
	White	65	.68		64
Genuine Persuasive	African	169	.67	1.04	168
	White	65	.68		64
Reserved Gregarious	African	169	.71	1.11	168
	White	65	.74		64
Cynical Trusting	African	169	.77	1.36	168
	White	65	.83		64
Optimistic Pessimistic	African	169	.71	.96	168
	White	65	.70		64
Emotional Phlegmatic	African	169	.61	1.25	168
	White	65	.68		64
Accommodating Assertive	African	169	.57	1.83*	168
	White	65	.77		64
Composed Contesting	African	169	.78	.73	168
	White	65	.70		64
Abstract Pragmatic	African	169	.49	1.84*	168
	White	65	.72		64
Motivational Distortion	African	169	.61	1.14	168
	White	65	.66		64

* *p* < .05

For the Abstract Pragmatic scale, there were no significant differences between the race groups.

For the Motivational Distortion scale, significant differences were found between the mean of the African respondents ($M = 22.78$) and the mean of the white respondents ($M = 24.26$), with $t(116) = 2.40, p = .018; d = .35$ ($SD = .15, df = 232$) and the 95% confidence limits are $.06 - .64$.

Reliability coefficients for African and white respondents

The reliability of the OPP for the race groups was determined by means of reliability analysis. Cronbach's coefficient alpha was computed in order to establish the reliabilities of the African and white sub samples. The results are reflected in Table 5, where the reliability coefficients of the African and white respondents are compared. To test equality of the two independent reliability coefficients, the statistic $F = (1 - \alpha_1)/(1 - \alpha_2)$ was used, in which α_1 and α_2 represent the reliabilities of the African and white respondents, respectively. The statistic followed an F distribution with $N_1 - 1$ and $N_2 - 1$ degrees of freedom (Van de Vijver & Leung, 1997).

Reliability studies could not be carried out for the language groups as some subgroups were too small to provide meaningful statistical comparisons.

The highest reliabilities for both groups were found on the Detail Conscious Flexible, Genuine Persuasive, Reserved Gregarious, Cynical Trusting, Optimistic Pessimistic, Emotional Phlegmatic, Composed Contesting and Motivational Distortion scales. There were however significant differences between the reliability coefficients of the African and white respondents on the Accommodating Assertive $F(168, 64) = 1.83, p < .05$ and the Abstract Pragmatic scales $F(168, 64) = 1.84, p < .05$. For both scales, the reliability was significantly higher for the white respondents.

DISCUSSION

When considering the results, research suggests that the most common factors that need to be considered as potential sources for item and/or test bias are race and language (De Beer, 2004). Abrahams (2002) contends that the race variable tends to have the greatest influence on test scores, whilst Meiring (2007) reports that language was the major factor impacting on test bias. Based on the results of this study, significant differences in mean personality trait scores among various race and language groups were found, as consistent with previous research (*cf.* Abrahams, 1996; Abrahams & Mauer, 1999; Packman *et al.*, 2005; Ones & Anderson, 2002; Ones & Viswesvaran, 1998). In analysing mean differences between groups, effect sizes (d values) will be considered as they provide an indication of possible adverse impact, i.e., the probability of selecting a candidate from a particular group is unacceptably low in comparison with other groups (Aamodt, 1999). Only medium to large effect sizes will be considered in the present discussion.

Differences between groups

Of the nine scales that make up the OPP, significant differences among groups were found on seven scales, all showing medium to large effect sizes.

For the Detail Conscious Flexible scale, the mean score for the English language group was significantly lower than the mean of the African language group ($d = .53$, indicating a medium effect size). Based on the OPP scale descriptions, this suggests that the English language group on average is more meticulous, rule bound, rigid, and conscientious, while the African language group is more unconventional, flexible, and lacking in self discipline.

The mean score of the white respondents was also significantly lower than the mean of the African respondents ($d = .68$, indicating a medium effect size) on the Detail Conscious Flexible scale. Similarly, this implies that the white respondents tend to be more meticulous, rule bound, rigid, and conscientious, while the African group is more unconventional, flexible, and lacking in self discipline.

For the Genuine Persuasive scale, the mean score of the English language group was signifi

cantly lower than the mean of the Afrikaans language group ($d = .96$, indicating a large effect size). This implies that the English language group is more forthright, genuine, and sincere, while the Afrikaans language group is more diplomatic, persuasive, and socially aware.

For the Cynical Trusting scale, the mean score of the English language group was significantly lower than the mean of the African language group, ($d = .67$, indicating a medium effect size). The English language group therefore has a tendency to be more suspicious, skeptical, and cautious in their approach, while the African language group is more trusting and somewhat naïve.

For the Optimistic Pessimistic scale, the mean score of the African language group was significantly lower than the mean score of the English language group, ($d = .63$, indicating a medium effect size). This implies that the African language group is more positive and achievement orientated, believing in their own abilities, while the English language group is more negative and resigned, with a tendency to doubt their own abilities to overcome problems.

A significant difference between means on the Optimistic Pessimistic scale was also found for the race groups. The mean score of the African respondents was significantly lower than that of the white respondents, ($d = .51$, indicating a medium effect size). Correspondingly, this implies that the African respondents are more positive and achievement orientated, believing in their own abilities, while the white group is more negative and resigned, with a tendency to doubt their own abilities to overcome problems.

For the Accommodating Assertive scale, the mean score of the African language group was significantly lower than the mean of the Afrikaans language group, ($d = .63$, indicating a medium effect size). The African language group is therefore more accepting, modest, and unassuming, while the Afrikaans language group is more dominant, challenging, and confrontative.

For the Composed Contesting scale, the mean score of the Afrikaans language group was significantly lower than that of the African language group, ($d = .60$, indicating a medium effect size). This implies that the Afrikaans language group has a calmer, more tolerant approach, preferring to work at a steady pace, while the African language group is more ambitious, competitive, and prone to worry.

Similarly, the mean score of the white respondents was significantly lower than the mean of the African respondents, ($d = .67$, indicating a medium effect size) on the Composed Contesting scale, suggesting that the white respondents have a calmer, more tolerant approach, preferring to work at a steady pace, while the African group is more ambitious, competitive, and prone to worry.

For the Abstract Pragmatic scale, the mean score of the Afrikaans language group was significantly lower than the means of the African and English language groups, ($d = .86$ and $.79$, respectively, indicating large effect sizes). This implies that the Afrikaans language group is more imaginative and creative, while the African and English language groups are more concrete, practical, and realistic.

In terms of reliability, the OPP shows acceptable alpha coefficients on a number of scales. However, the internal consistencies of the Accommodating Assertive (mean alpha $= .57$) and Abstract Pragmatic (mean alpha $= .49$) scales for the African group are significantly lower than that of the white group.

Practical implications

In the present work we are analysing differences among groups on a subscale level because, in our experience, the subscales of personality tests are used to measure specified competencies during selection, placement, and development of employees. For example, the Detail Conscious Flexible subscale may be used to measure a competency such as “detail orientation” and in accordance with the EE Act this competency would be linked to an “inherent requirement of the job”. The subscale scores on the personality test would then be combined with other measurements, for example, a role

play, case study, or leaderless group discussion, of the same competency and eventually the combined score would be used in the decision making process. As is evident from the results, the mean score of the African language group is significantly higher than that of the English language group ($d = .53$). As a result of his/her score on the Detail Conscious Flexible subscale, it may happen that an African language speaker's combined score on the competency under discussion exceeds a cut off score and consequently his/her application is turned down.

It should however be noted that significant differences in test scores do not necessarily imply adverse impact. The results of a personality test will typically result in adverse impact if individuals from a specific group are less likely to be selected for employment opportunities. Differences in responses on personality tests are relevant to the extent that they influence selection outcomes within organisations (Ones & Anderson, 2002). If it is known that such differences exist, the standard error of measurement of the test may be taken into consideration or the contribution of the particular subscale may be weighted when interpreting the combined scores of the individuals.

Test users should nonetheless be mindful of how the results of psychological tests are used and subsequent decisions are made in employment settings. Based on the results of the present study, it is important to note that psychological tests should never be used on their own for making decisions about placement or promotion, but should rather be used as an aid in decision making. The consequences of using one test only can be devastating, not only for the individual being tested, but also for the organisation in terms of costly lawsuits.

In terms of the broader assessment context, the EEA has called for the development of more culturally appropriate psychological tests and the validation of existing tests for use in diverse groups. There is evidence to suggest that developing South African tests is the most viable option (Meiring, 2007). Despite this need, very few new measures have been developed locally (Meiring, 2007; Taylor & De Bruin, 2005). Foxcroft *et al.* (2004) listed key aspects to guide the continued development of psychological testing in South Africa. On the agenda is the development of training programmes to build and increase test development expertise locally. Until this has been achieved, many are in favour of adapting and standardising well researched international tests. This does however also present challenges as many adapted versions of imported tests have been proved to be non comparable within the South African context.

In addition, almost all psychological tests available for use in South Africa are in English. Meiring (2007) and Saunders (2002) cautions against the use of English as a medium of assessment, considering that the majority of the South African population are not adequately fluent in English. Only 10% of the South African population speak English as a home language, of which two in three are not white (International Marketing Council of South Africa, 2006). The 2001 census further revealed that English is the mother tongue of only 8.2% of the population (International Marketing Council of South Africa, 2006). It would therefore seem unfair to use English as a medium of assessment. There is a tendency to assume that if a person speaks English, they should be able to understand the level of English in a personality test (Saunders, 2002). There is therefore the possibility that these individuals may be significantly disadvantaged when tested in English, which may amount to unfair discrimination. Tredoux (n.d.) therefore suggests assessing an individual's English proficiency to determine its impact on performance in personality tests. Prinsloo and Ebersöhn (2002) also support testing an individual's performance on a personality test according to their English proficiency. Assessing an individual's ability to not only speak, but read and write, in English can assist in making relatively accurate judgments about the degree to which language disadvantaged the individual (Saunders, 2002).

In addition to this, Van Der Merwe (1999) suggests validating psychological tests for organisations in their own work environments and creating organisation specific norm groups. Van Der Merwe (1999) also notes that validating test batteries for specific jobs could further enhance the

fairness in assessment processes within organisations. He also recommends making more use of competency based exercises which are directly related to job content and inherent requirements. Saunders (2002) recommends the use of behaviour observation or other similar means of obtaining information about performance. She questions the use of personality testing in any diverse setting, where issues such as language or culture are likely to be prevalent. Considering the diversity of our society, this is likely to be the case in many South African organisations. The value of personality testing is not disputed but rather its applicability in certain contexts. Based on the results of the present study, one cannot rule out the possibility that cultural variables could have impacted on the scores on the OPP.

LIMITATIONS AND CONCLUSION

There are several limitations of the present study that require consideration. It should firstly be highlighted that Tredoux (n.d.) clearly provides the requirements for effective completion of the OPP, which includes a Grade 12 level of education as well as proficiency in English. A critical limitation of this study is therefore that the respondents within the sample were not pre screened for English, which may have negatively impacted on the value of the results. An assumption was made that as the majority of the respondents are already functioning in high level management roles, English proficiency can be assumed. Evidently, language may have a significant impact on test performance. An additional and obvious limitation is that the sample was relatively small. Caution regarding the interpretation of the results is therefore recommended. This is particularly true for any findings involving the white sample, which was very small ($n = 65$).

As it was the intention from the beginning of this study to analyse results on a subscale level because these scores are used in decision making, individual items were not analysed. A suggestion is to do Differential Item Analysis in a further study, to determine whether group differences could not be explained in terms of individual items.

In conclusion, the results of the present study have allowed for an estimate of group differences on the OPP used for selection purposes. Although the sample was relatively small, these findings may provide valuable information in terms of whether or not the OPP is suitable for use across different racial and language groups, which may in turn provide assistance in determining the value of this test for making occupationally related decisions. According to Van de Vijver and Leung (1997), an investigation into the psychometric adequacy of the OPP is an important initial step in analysing cross cultural applicability. Further research is required to determine the exact causes of these differences.

REFERENCES

- Aamodt, M.G. (1999). *Applied Industrial/Organizational Psychology* (3rd edn). Belmont, CA: Cole Wadsworth.
- Abrahams, F. (1996). The cross cultural comparability of the 16PF. Unpublished doctoral thesis, Department of Industrial Psychology, University of South Africa.
- Abrahams, F. (2002). The (un)fair usage of the 16PF (SA92) in South Africa: a response to C.H. Prinsloo and I. Ebersöhn. *South African Journal of Psychology*, 32, 58-61.
- Abrahams, F. & Mauer, K. F. (1999). The comparability of the constructs of the 16PF in the South African context. *Journal of Industrial Psychology*, 25, 53-59.
- Barrick, M.R. (2001, June). *Personality testing, controversial no more*. Keynote address presented at the fourth annual Industrial Psychology conference, Pretoria, South Africa. Retrieved August, 3, 2005, from http://www.siopsa.org.za/Conference/2001archive_keynote.htm
- Barrick, M. R. & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta analysis. *Personnel Psychology*, 44. Retrieved August, 3, 2005, from <http://weblinks1.epnet.com>
- Barrick, M. R., Mount, M. K. & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9, 9-30.

- Bedell, B., Van Eeden, R. & Van Staden, F. (1999). Culture as moderator variable in psychological test performance: Issues and trends in South Africa. *South African Journal of Industrial Psychology*, 25, 1-7.
- Chamorro Premuzic, T. & Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality*, 37, 319.
- De Beer, M. (2004). Use of differential item functioning (DIF) analysis for bias analysis in test construction. *South African Journal of Industrial Psychology*, 30, 52-58.
- Employment Equity Act, Act 55, 1998. *Republic of South Africa Government Gazette*, 400 § 19370.
- Foxcroft, C. (2004). Planning a psychological test in the multicultural South African context. *South African Journal of Industrial Psychology*, 30, 8-15.
- Foxcroft, C., Paterson, H., le Roux, N. & Herbst, D. (2004). *Psychological assessment in South Africa: a needs analysis. The test use patterns and needs of psychological assessment practitioners*. Pretoria: The Health Professions Council of South Africa.
- Goldberg, L.R., Sweeney, D., Merenda P.F., & Hughes, J.E., Jr. (1998). Demographic variables and personality: The effects of gender, age, education, and ethnic/racial status on self descriptions of personality attributes. *Personality and Individual Differences*, 24, 393-403.
- International Marketing Council. (2006). *The languages of South Africa*. Retrieved May, 15, 2007, from http://www.southafrica.info/ess_info/sa_glance/demographics/language.htm
- Johnson, D. L. & Kleiner, B. H. (2000). A comparative view of employment testing. *Management Research News*, 23, 99-102.
- La Grange, L. & Roodt, G. (2001). Personality and cognitive ability as predictors of the job performance of insurance sales people. *Journal of Industrial Psychology*, 27, 35-43.
- Lievens, F., Coetsier, P., De Fruyt, F. & De Maeseneer, J. (2002). Medical students' personality characteristics and academic performance: A five factor model perspective. *Medical Education*, 36, 1050-1056.
- Meiring, D. (2000, June). *Revisiting the cross cultural comparability of the 16 Personality Factor Inventory (16PF) in the South African context*. Paper presented at the Industrial Psychology conference (incorporating the Psychometrics conference), Pretoria, South Africa.
- Meiring, D. (2007). Bias and equivalence of psychological measures in South Africa. Unpublished doctoral thesis, University of Tilburg.
- Meiring, D., Van de Vijver, F. J. R. & Rothmann, S. (2006). Bias in an adapted version of the 15FQ+ Questionnaire in South Africa. *South African Journal of Psychology*, 36, 340-356.
- Meiring, D., Van de Vijver, F. J. R., Rothmann, S. & Barrick, M. R. (2005). Construct, item and method bias of cognitive and personality tests in South Africa. *South African Journal of Industrial Psychology*, 31, 1-8.
- Ones, D. S. & Anderson, N. (2002). Gender and ethnic group differences on personality scales in selection: Some British data. *Journal of Occupational & Organisational Psychology*, 75, 255-276.
- Ones, D.S., & Viswesvaran, C. (1998). Gender, age and race differences on overt integrity tests: Analyses across four large scale applicant data sets. *Journal of Applied Psychology*, 83, 35-42.
- Owen, K. (1986). *Test and item bias: Administration of the Senior Aptitude Test, Mechanical Insight Test, and the Scholastic Proficiency Battery to white, Indian, black and coloured Technikon students*. Pretoria: Human Sciences Research Council.
- Packman, T., Brown, G., Englert, P., Sisarich, H. & Bauer, F. (2005). Differences in personality traits across ethnic groups in New Zealand and across an international sample. *New Zealand Journal of Psychology*, 34, 77-85.
- Prinsloo, C. H. & Ebersöhn, I. (2002). Fair usage of the 16PF in personality assessment in South Africa: A response to Abrahams and Mauer with special reference to issues of research methodology. *South African Journal of Psychology*, 32, 48-57.
- Psytech International. (n.d.). *Occupational Personality Profile: The technical manual*. Retrieved February, 26, 2005, from <http://www.psytech.co.uk/downloads/manuals/oppman.pdf>
- Retief, A.I. (1992). The cross cultural utility of the SAPQ – bias or fruitful differences? *South African Journal of Psychology*, 22, 202-207.
- Rosenthal, R., & Rosnow, R.L. (2008). *Essentials of Behavioral Research: Methods and Data Analysis* (3rd edn). Boston: McGraw Hill.

- Salgado, J.F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology, 82*, 30-45.
- Saunders, E. (2002). *Assessing human competence: Practical guidelines for South African managers*. Randburg, South Africa: Knowledge Resources.
- Schmidt, F.L., Ones, D.S., & Hunter, J.E. (1992). Personnel selection. *Annual Review of Psychology, 43*, 627-670.
- Taylor, T.R., & Boeyens, J.C.A. (1991). The comparability of the scores of blacks and whites on the South African Personality Questionnaire: an exploratory study. *South African Journal of Psychology, 21*, 1-11.
- Taylor, N., & De Bruin, G. P. (2005). *Basic Traits Inventory: Technical manual*. Johannesburg: Jopie van Rooyen & Partners SA.
- Tredoux, N. (n.d.). *Occupational Personality Profile: South African users guide and research reference*. Johannesburg: Psytech SA.
- Van de Vijver, F. & Leung, K. (1997). *Methods and data analysis for cross cultural research*. London: Sage Publications.
- Van Der Merwe, R. P. (1999). Psychological assessment in industry. *South African Journal of Industrial Psychology, 25*, 8-11.
- Van der Walt, H. S., Meiring, D., Rothmann, S. & Barrick, M. R. (2002). *Meta analysis of the relationship between personality measurements and job performance in South Africa*. Paper presented at the fifth annual conference of the society for Industrial & Organisational Psychology of South Africa, Pretoria, South Africa. Retrieved August, 3, 2005, from <http://www.siopsa.org.za>.