

Theoretical and methodological considerations in the translation of the 16PF5 into an African language



René van Eeden

Department of Psychology, University of South Africa, P. O. Box 392, Pretoria, 0003, South Africa
veeder@unisa.ac.za

Tshifhiwa Rebecca Mantsha

Centre for Student Counselling, Social and Academic Development, University of Venda for Science and Technology, Thohoyandou, South Africa

A Tshivenda version of the Sixteen Personality Factor Questionnaire Fifth Edition (16PF5) was developed using a back-translation design. The translated version was administered to a Tshivenda-speaking sample. Analyses were done at an item level, specifically in terms of the contribution of the items to the reliability of the 16 primary factors. Results indicated that many of the items would have to be excluded if the aim were to achieve acceptable reliability, and even then the reliability coefficients would remain low. In the case of a number of the ineffective items, translation changed the meaning of the items because of the absence of an equivalent concept in the target language, difficulty in translating idiomatic expressions, potential confusion due to the use of the negative form, and translation errors. Trends that indicate cultural differences in the manifestation of constructs were also observed and these were related to cultural norms (e.g., the expression of emotions) and experiential factors (e.g., job-related experience). A literal translation of the questionnaire proved to be insufficient and the findings of this study need to be considered in further adaptations of the test. The theoretical and methodological issues highlighted here have implications for the translation of this questionnaire into other African languages as well as for the adaptation of personality questionnaires in general.

Keywords: 16PF; 16PF5; back-translation; cross-cultural adaptation; personality assessment; reliability; Tshivenda translation; Venda translation

The influence of South Africa's political, economic, and social history on psychological testing practices has been explored by authors such as Claassen (1997) and Foxcroft (1997). Changes in these spheres have resulted in pressure on test developers and test users to develop and use culturally appropriate assessment instruments and practices. However, test development and adaptation in a multicultural, multilingual society in transition is

a complex process, and progress in this regard has been slower than expected (Foxcroft, 1997). South African society is heterogeneous in terms of factors that are considered to moderate performance on psychological tests. Variables such as language proficiency, culture, education, socio-economic status, home environment, urbanisation, and test-wiseness have been identified (Grieve, 2005; Nell, 1997; Owen, 1996). Research on cognitive tests and personality tests justifies the need for considering these factors when developing, adapting, and using psychological tests in a multicultural context (Abrahams & Mauer, 1999a, 1999b; Claassen & Hugo, 1993; Claassen, Krynauw, Holtzhausen, & Wa ga Mathe, 2001; Claassen & Schepers, 1990; Meiring, Van de Vijver, & Rothmann, 2006; Meiring, Van de Vijver, Rothmann, & Barrick, 2005; Owen, 1991, 1992; Prinsloo, 1992, 1998; Shuttleworth-Edwards et al., 2004; Taylor & Boeyens, 1991; Tollman & Msengana, 1990; Van Eeden, 1993; Van Eeden & Prinsloo, 1997; Van Eeden, Taylor, & Du Toit, 1996; Van Eeden & Van Tonder, 1995; Van Eeden & Visser, 1992). This article explores the role of language and culture in adapting the Sixteen Personality Factor Questionnaire Fifth Edition (16PF5) for the Venda population.

Language proficiency as a potential source of bias is frequently mentioned in relation to ability testing (Claassen & Hugo; 1993; Claassen et al., 2001; Foxcroft, 1997; Grieve, 2005; Nell, 1994; Owen, 1991; Van Eeden, 1993; Van Eeden & Van Tonder, 1995), but the same arguments apply to personality questionnaires (Maree, 2002; Prinsloo, 1998; Van Eeden et al., 1996). Van den Berg (1996) states that, when there is a difference between home language and language of instruction, testing in either language puts the testee at a disadvantage. Bilingual assessment is suggested as a possible solution (Foxcroft, 1997; Van den Berg, 1996), while the consideration of language proficiency is advised in the case of cross-cultural studies and to enhance fair testing practice (Claassen et al., 2001; Claassen & Hugo, 1993; Foxcroft, 1997; Grieve & Van Eeden, 1997; Maree, 2002; Meiring, Van de Vijver, & Rothmann, 2006; Meiring et al., 2005; Owen, 1991; Prinsloo, 1998; Prinsloo & Ebersöhn, 2002; Van Eeden, 1993; Van Eeden et al., 1996; Van Eeden & Van Tonder, 1995). Practical problems for translation in the South African context include the large number of official languages and the availability of test administrators who speak these languages. Practitioners also report different dialects and a lack of language standardisation in the case of African languages, and Wallis and Birt (2003) point out that languages evolve over time. These languages also often lack the concepts and expressions required for equivalence in translation (Grieve, 2005; Van Den Berg, 1996).

Wallis and Birt (2003) refer to the unique role that culture plays in the case of personality tests, due to its influence on individuals' affective processes. Research on the use of personality questionnaires based on Western concepts of personality (such as the 16PF) in culturally heterogeneous environments should explore the etic (universal) or emic (culture-specific) nature of the constructs measured by such tests. This is especially relevant to trait theory-based instruments where behaviour is assumed to be influenced by a number of underlying traits or dimensions. Laher (2007) argues that African and Asian perspectives on personality differ vastly from Western perspectives. Not only can

different manifestations of personality dimensions be expected (e.g., collectivism versus individualism, sociocentric versus egocentric, and interdependence versus independence), but indigenous constructs not accounted for by Western models of personality have also been identified (e.g., the interpersonal relatedness factor identified in a Chinese population, and a spiritual element that forms part of the Asian perspective).

Various authors (e.g., Marsella, Dubanoski, Hamada, & Morse, 2000; Stumpf, 1993; Wilson, Doolabh, Cooney, Khalpey, & Siddiqui, 1990) refer to the generalisation of major personality dimensions across cultural boundaries but caution that the cultures may differ in terms of their position on these dimensions. Robust constructs defined in broad terms can be expected to show comparability across groups. The five factor model of personality has, for example, been replicated in some local cross-cultural studies (De Bruin, Schepers, & Taylor, 2005; Heuchert, Parker, Stumpf, & Myburgh, 2000) although differences in this regard have also been found between black and white samples (De Bruin, 2005; Taylor, 2000). Differences in the meaning attached to constructs and in the way constructs manifest are expected, given the highly socialised nature of the concepts measured by personality tests. Laher (2007) regards personality as a dynamic concept due to its interaction with the environment. Van de Vijver and Poortinga (1992) refer to the intricate relationship between culture and test performance, and, according to Retief (1992), analysing differences found on measures of personality can help explain behavioural rules and attributions made in certain situations by different cultures. Existing tests can thus be used in the South African context to formulate 'a theory of personality trait manifestation, in terms of behaviour preferences as moderated by social and cultural context' (Retief, 1992, p. 206). Meiring, Van de Vijver, De Bruin, et al. (2006) also refer to the Chinese example of accounting for culture specifics in an indigenous personality psychology and the need for a similar development in the South African context.

South African research on the equivalence for different language and population groups of the English-language versions of the 16PF showed mixed results. Prinsloo (1992) reported favourable results when comparing the difficulty values (and their ranking) and the discrimination values of the items of the 16PF South African 1992-version (SA92). No significant differences were found in terms of the reliability coefficients, and the values for both the primary and the second-order factors were acceptable. Van Eeden and Prinsloo (1997) found numerically lower reliability coefficients in their study with the 16PF (SA92), but the values were mostly acceptable. Similar second-order factor structures were reported for the different groups in both studies, but some culture-specific trends were noted that have to be taken into account when interpreting scores on these factors (Prinsloo, 1992; Van Eeden & Prinsloo, 1997). Abrahams and Mauer (1999a, 1999b) found less favourable results for the 16PF (SA92). Approximately one fifth of the items on this form did not attain satisfactorily item-total correlations for a black sample, and differential item analysis also showed substantial differences between the ways in which black and white respondents reacted to the items. The coefficients of internal consistency on the primary factors were also unacceptably low for the black sample. In addition,

consistent differences have been found in terms of the average performance on a number of factors when comparing white and black samples in research with the 16PF (SA92) (Abrahams & Mauer, 1999a; Prinsloo, 1992; Van Eeden & Prinsloo, 1997).

In the case of the 16PF5, Van Eeden et al. (1996) found that, although results in terms of differential item functioning were reasonable, the item-total correlations of a large number of items were inadequate for African language speakers. Numerically lower values were also found for this group in terms of the reliability of the primary factors, with some of the reliability coefficients being unacceptably low. Factorial similarity could also not be shown for the different groups. De Bruin et al. (2005), however, reported similarity in terms of the second-order factors for Afrikaans, English, and African-language speakers. Differential item analysis and factor analysis also produced reasonably favourable results for the 16PF5 when controlling for language proficiency (Prinsloo, 1998). Differences were found in the average profiles of white and black samples (Maree, 2002; Van Eeden et al., 1996), and similar results were also found when comparing groups differing in language proficiency (Maree, 2002).

Differential item functioning and different response rates to items (resulting in differences in mean profiles) could reflect real differences in terms of the constructs being measured by a test (Prinsloo & Ebersöhn, 2002; Retief, 1992). This, however, refers specifically to differences that correspond with theory-based and real-life differences in the underlying trait or traits, and one needs to distinguish between useful differences and potential bias and discrimination (Van de Vijver & Tanzer, 1997). Reliability coefficients should furthermore be acceptable, and, as construct comparability is central to cross-cultural personality studies, equivalence at this level is non-negotiable. Given the findings of the research with the English-language versions of the 16PF, adaptation of the versions has to be considered.

Reasons for cross-cultural adaptation of psychological tests include the fairness implied by testing in a language of choice, facilitation of comparative studies, and savings in terms of costs and in terms of time (Hambleton & Kanjee, 1995). According to Ralston (1995), individuals also respond in terms of the culture of the language of presentation and, by using non native-language instruments, valuable cross-cultural information is lost. The term 'adaptation' is preferred, as translation is part of a broader adaptation process that provides for linguistic as well as cultural factors (Hambleton, 1994).

Van de Vijver and Tanzer (1997) describe three options for adapting an instrument to ensure that language and culture are taken into account. These options imply change in an increasing number of items. The first option, application, is based on the assumption that the underlying constructs are appropriate in the different cultural groups and that a literal translation of the instrument into the target language will therefore be sufficient. The second option is adaptation, which involves a combination of literal translation, changes in items, and the development of new items. This option ensures that culturally idiosyncratic expressions of the constructs are included. For the third option, assembly, the level of change implies that in effect a new instrument is assembled. This option is

used when there is differential appropriateness of the item content for the majority of the items, if there is incomplete overlap of the construct definitions (implying that all aspects of a construct are not covered), or if indigenous personality constructs are not covered.

The literal translation option is the most common, as it allows for direct comparison on the different language versions (Van de Vijver & Tanzer, 1997). Requirements for the selection of translators include competence in the relevant languages, knowledge of the cultures involved, domain knowledge, and psychometric experience. The translation–back-translation procedure, which has been widely applied in multicultural contexts (Bracken & Barona, 1991; Hambleton & Kanjee, 1995; Kanjee, 2005; Van de Vijver & Tanzer, 1997; Van Ede, 1996), involves the translation of the source version of the test into the target language, followed by translation back into the original language by a different translator or translators who have no knowledge of the original test. Source language experts then evaluate the original and back-translated versions for equivalence.

For the present study, a literal translation of the 16PF5 into Tshivenda was regarded as a logical first step before considering more complex forms of adaptation. The back-translation design was feasible from a practical point of view (rather than, for example, the committee approach) and this design was also used to develop a Spanish version of the 16PF5 (Ellis, 1995). The aim was to determine whether the translated test was psychometrically sound, to identify possible language and cultural issues that needed to be considered if further adaptation was necessary, and to inform decisions regarding equivalence studies.

METHOD

The present study was designed to assess the psychometric integrity of the Tshivenda translation of the 16PF5 for target language monolinguals. It was assumed that the respondents were most competent in their mother tongue even if they were probably also speakers of the source language (English) and other languages. This assumption was, however, not tested and might be somewhat simplistic, given Van den Berg's (1996) view on the role of both the home language and the language of instruction.

Sample

An opportunity was negotiated to apply the Tshivenda translation of the 16PF5 to the first-, second-, and third-year Industrial Psychology and Psychology students at the University of Venda for Science and Technology. The students were competent in Tshivenda and most of them shared the cultural origins of Tshivenda as mother tongue. The rural context also implies greater identification with traditional Venda culture than in the case of university students in an urban context where greater exposure to Western culture might mean that students are more likely to adopt aspects of the latter (Marsella et al., 2000). Relatively uniform socio-economic conditions were also expected. The students were chosen from the

specific departments to limit extraneous variables in terms of interest (and possibly career aspirations), and it was also expected that the language skills of the students reading either of the subjects would be similar. Prinsloo and Ebersöhn (2002) discuss the limitations in terms of the generalisability of results when using undergraduate psychology students. However, given the exploratory nature of the study, it was decided that a homogeneous group at this stage would simplify the interpretation of the findings. The sample consisted of 85 students with 92 per cent having Tshivenda as first language. The distribution of the sample was equal in terms of both gender and study level (i.e., first, second, and third year), while the age of the students ranged from 18 years to 39 years, with the majority (78 per cent) being in their twenties.

Measuring instrument

The 16PF5 was released in the United States in 1994 and is an updated and revised version of earlier forms of the 16PF (Conn & Rieke, 1994). The intention was to select and update the best items from current forms of the test and to combine these with new items. The developers considered the following issues during item selection: the item-scale correlations; item frequency endorsements; simplicity of the language; updating the language; possible gender, race, and disability bias; content that is cross-culturally translatable; avoiding content that is socially desirable or undesirable; and avoiding material that might be considered unacceptable in an industrial or organisational setting.

The 16PF5 comprises 185 items. In addition, 64 items were developed for research purposes, such as possible replacement items in cross-cultural studies. These items are referred to as Form S. Each item consists of a statement with three response options scored as 0, 1, or 2. The b responses to personality items appear as a question mark, thus providing a uniform response choice that covers several different reasons for not selecting either the a or c alternative.

The items are combined into 16 primary personality traits, named in the manual (Russell & Karol, 1994) as Warmth, Reasoning, Emotional stability, Dominance, Liveliness, Rule-consciousness, Social boldness, Sensitivity, Vigilance, Abstractedness, Privatness, Apprehension, Openness to change, Self-reliance, Perfectionism, and Tension. These traits are combined in groups to obtain scores on six second-order factors, namely Extraversion, Anxiety, Self-control, Independence, Tough-mindedness, and a loading on the primary factor B. Faking scales were also developed. Support for the reliability and validity of the American version is discussed in the manual.

The American versions of the 16PF5 and of Form S were converted by the Human Sciences Research Council (HSRC) into South African English with minimal adaptations (Prinsloo, 1998; Van Eeden et al., 1996). Slight changes, often in the form of explanatory phrases, were made with consideration of input from the developers. In the present study, this adapted version was translated into Tshivenda using the back-translation design. The English version was translated into Tshivenda by a translator, and the Tshivenda version

was translated back into English by a different translator who had no knowledge of the original test. Both translators are Tshivenda speakers who are fluent in Tshivenda and English and who were, at the time of the study, employed in language departments at the University of Venda for Science and Technology. The first author of this article, who is an expert in the field of psychometrics, personality testing, and specifically the 16PF, compared the original English version and the back-translated English version in terms of technical details, language usage, the meaning conveyed, and the constructs measured. Approximately a quarter of the items were problematic. Examples include the following phrases: ‘lively party’ in the original version and ‘good feast’ in the back-translated version, and ‘articles on today’s social problems’ versus ‘a piece of recent problems’. The first translator reconsidered the Tshivenda translation of the problematic items and the improved items were translated into English by the second translator. These items were again compared across the two English versions by the first author together with an expert from the HSRC. The few items that were still problematic in terms of the construct being measured were corrected in the Tshivenda version by the second author of this article, a Tshivenda speaker.

Procedure

The Tshivenda version of the 16PF5 was administered to second- and third-year Psychology students and first-, second-, and third-year Industrial Psychology students at the University of Venda for Science and Technology. The administration took place during two sessions with students from the different departments participating in different sessions. The aim of the study was explained and students were assured of the confidentiality of the results. Mention was made of the value of personality questionnaires and of the need for cross-cultural research in this regard. Students were informed that the results would be used at a group level and not at an individual level. Although participation was voluntary, students might have been reluctant not to participate given the formal setting and possible peer pressure.

Data analysis

After data capturing and editing of the data, 85 complete records remained. Analysis commenced at an item level, specifically in terms of the contribution of the items to the reliability of the 16 primary factors. The means and standard deviations of the items (including the additional Form S items) of each primary factor were calculated. Each of the original items of a factor (excluding the additional items) was thereafter correlated with the corrected total factor score (the total with the item in question excluded from the total). Items with a negative item-total correlation were excluded and the analysis was redone with the remaining items. This was repeated until all items correlated positively with the total. The effect that the exclusion of an item would have on Cronbach’s alpha value for the

factor was considered next and, if there was a considerable increase in Cronbach's alpha, the item was excluded and the analysis was redone with the remaining items. This process was repeated until either no further increase in the internal consistency reliability could be achieved or at least five items remained. After this, the additional items were added and the process was repeated. When no further improvement could be achieved by excluding items, the items of the last run were considered to make up the factor.

RESULTS

An overall evaluation of the results indicated that many of the items had to be excluded if the aim were to achieve acceptable reliability (see Table 1). Even after the exclusion of items, the reliability coefficients remained low with values above 0.60 being obtained for only four of the factors. To illustrate the data analysis process, detailed results are presented for Factor A (Warmth) (one of the factors with a large number of problematic items). Following this, a summary of the findings for each factor is given. A detailed theoretical interpretation of the items to be excluded from each factor would probably have resulted in capitalising on chance and an over-interpretation. Only clearly identifiable trends are therefore mentioned.

Table 1. Results of the reliability analysis

Factor	Original items		Additional items	
	Number	Accepted	Number	Accepted
Factor A: Warmth	11	5	5	2
Factor B: Reasoning	15	10	3	2
Factor C: Emotional Stability	10	4	7	6
Factor E: Dominance	10	7	4	1
Factor F: Liveliness	10	9	4	1
Factor G: Rule-consciousness	11	9	3	1
Factor H: Social Boldness	10	9	4	2
Factor I: Sensitivity	11	4	3	3
Factor L: Vigilance	10	5	4	2
Factor M: Abstractedness	11	8	5	5
Factor N: Privateness	10	7	4	3
Factor O: Apprehension	10	10	5	-
Factor Q1: Openness to Change	14	11	-	-
Factor Q2: Self-reliance	10	10	4	4
Factor Q3: Perfectionism	10	6	5	5
Factor Q4: Tension	10	10	4	-
Total	173*	124	55 (64)**	37

Notes: * The Impression Management Scale, consisting of 12 items, was not included in the analysis. ** In the case of Factors O and Q4, the additional items were not analysed.

Factor A (Warmth)

Table 2 includes the means and standard deviations of all the items of Factor A; that is, the original as well as the additional items. The means indicate that this sample answered in a 'positive' direction on approximately two thirds of the items with average values for most of the rest of the items. In terms of variability, however, it is clear that responses varied across a broad range, thus showing potential for discrimination between respondents in terms of this trait.

Table 2. Means and standard deviations of the items of Factor A

Items	Mean	Standard deviation
A1	1.75	0.653
A2	1.58	0.730
A3	1.73	0.625
A4	0.94	0.956
A5	0.94	0.943
A6	1.79	0.465
A7	1.74	0.601
A8	1.16	0.962
A9	1.68	0.727
A10	1.66	0.716
A11	0.92	0.954
Additional 1	1.73	0.662
Additional 2	0.89	0.873
Additional 3	0.82	0.953
Additional 4	1.31	0.887
Additional 5	1.01	0.893

The results of the reliability analysis are presented in Table 3 with initial analyses including only the original items, following which the additional items were added and the process repeated. Five of the original 11 items and two of the additional five items remained in the final run. This factor measures a person's interest in people, his or her preference for participating, and warmth and caring towards others. Items deal with the expression of feelings and with an interest in specific jobs and preferences regarding job-related functions. Items dealing with feelings were mostly problematic, while some job-related items were also excluded. Potential confusion due to the use of the negative form was also noted.

Table 3. Sequential reliability analysis of the items of Factor A

Items	1st run		2nd run		3rd run		4th run		5th run	
	item-total correlation	Alpha item deleted	item-total correlation	Alpha item deleted	item-total correlation	Alpha item deleted	item-total correlation	Alpha item deleted	item-total correlation	Alpha item deleted
A1	0.2291	-0.0191	0.2995	0.3347	0.2557	0.3156	0.2462	0.3488	0.3346	0.3513
A2	-0.0894	0.1543	0.1656	0.4219	0.1242	0.3605	0.1337	0.3931	0.1058	0.4495
A3	0.1888	0.0076								
A4	-0.0159	0.1232								
A5	-0.0062	0.1153								
A6	-0.0865	0.1304								
A7	0.1034	0.0533	0.2879	0.3486	0.1840	0.3424	0.2388	0.3553	0.2734	0.3831
A8	-0.1673	0.2315								
A9	-0.0699	0.1435								
A10	0.1666	0.0076	0.1519	0.4353	0.1329	0.3570	0.1392	0.3916	0.1708	0.4245
A11	0.1854	-0.0397	0.2637	0.3615	0.3500	0.2439	0.3629	0.2602	0.3355	0.3256
Additional 1					0.1113	0.3646	0.1561	0.3845	0.1569	0.4297
Additional 2					0.1144	0.3662	0.1115	0.4103	0.1023	0.4694
Additional 3					0.1199	0.3659	0.0572	0.4459		
Additional 4					0.0755	0.3840				
Additional 5					0.0156	0.4106				
Alpha	0.0962		0.4364		0.3777		0.4087		0.4459	

Factor B (Reasoning)

Factor B is the only factor where there are correct answers and, based on the means, most of the items seemed difficult for this sample, although responses varied across a broad range. Ten of the original 15 items and two of the three additional items remained in the final run. There were no clear trends in terms of the problematic items, which deal with vocabulary, number and letter series, and knowledge. Difficulties in terms of translation were noted and a printing error was made in the case of one of the items.

Factor C (Emotional stability)

The means indicated that approximately half the items were answered in a ‘positive’ direction, with the rest of the values being average or below average. Responses again varied across a broad range. Unless otherwise indicated, this response pattern was similar for the rest of the factors. Only four of the original 10 items remained in the final run, but six of the seven additional items were found to be effective. Problematic items deal with variation in moods, emotional needs, and the ability to cope (although the items that were included also come from these categories). A translation error was noted, and in some cases translation might have changed the meaning. Examples of excluded items that were difficult to translate are: ‘I don’t let myself get depressed over little things’, and ‘When something upsets me, I usually get over it quite soon’. There is no Tshivenda expression for ‘depression’, and ‘upset’ was translated as ‘nndadisa’, which means ‘confused’. The use of the negative form possibly also resulted in confusion in certain instances.

Factor E (Dominance)

A strong ‘positive’ response pattern was observed. Seven of the 10 original items but only one of the four additional items remained in the final run. In the case of the problematic items, the focus was more on the interaction process (cooperation versus conflict) than on asserting oneself in terms of rules and standards. A translation error was noted and, in the case of the item ‘People regard me as a co-operative person; assertive/putting my point across strongly’, the explanation used could have changed the meaning of the item.

Factor F (Liveliness)

Nine of the original 10 items and one of the additional four items remained in the final run. Except for possible difficulty in translation (for example, ‘People think of me as a happy-go-lucky, carefree person’), no trends were observed.

Factor G (Rule-consciousness)

Nine of the 11 original items and one of the three additional items remained in the final run with no trends related to the problematic items.

Factor H (Social boldness)

A strong 'positive' response pattern was observed. Nine of the 10 original items and two of the additional four items remained in the final run. No trends were observed.

Factor I (Sensitivity)

In terms of the response pattern, no definite direction was observed for this factor, with means ranging evenly from below to above average. Four of the 11 original items and all three of the additional items remained in the final run. A choice between two activities or tasks related to activities is made on each of the items of this factor. In some instances, activities relate to gender stereotypes such as 'building', 'mechanic', or 'dancing', but this was not limited to problematic items and no trends were obvious in the case of the latter.

Factor L (Vigilance)

Five of the 10 original items and two of the four additional items remained in the final run. Translation might have changed the meaning of items such as: 'It's always important to pay attention to other people's motives', and 'It's wise to be on guard against smooth talkers because they might take advantage of you'. 'Motives' was translated as 'work', and 'smooth talkers' as 'sweet talk'. Possible confusion due to the use of the negative form was again noted.

Factor M (Abstractedness)

The response pattern showed no directionality for this factor. Eight of the 11 original items and all five additional items remained in the final run and no trends were observed.

Factor N (Privateness)

Again no directionality was observed in the response pattern. Seven of the 10 original items and three of the four additional items remained in the final run. A distinction can be made between items that imply sharing (problems, etc.) in a relatively close relationship versus being open with others in any situation. Problematic items primarily came from the category of items dealing with the latter. A translation error was noted.

Factor O (Apprehension)

A strong 'positive' response pattern was observed. All the original items were included in the first run and no further analysis was done with the additional items.

Factor Q1 (Openness to change)

Eleven of the 14 original items remained in the final run and there were no additional items for this factor. Other than difficulty in translation (for example, 'I am more interested in seeking personal meaning in life; a secure job that pays well' with 'personal meaning' translated as 'vhudipfi' which means 'feelings'), no trends were observed.

Factor Q2 (Self-reliance)

Means were below average for all but two items, indicating a tendency towards group affiliation. The 10 original and four additional items all remained in the final run.

Factor Q3 (Perfectionism)

A strong 'positive' response pattern was observed. Six of the 10 original items and all five of the additional items remained in the final run. Again no trends were observed, except possibly the use of the negative form.

Factor Q4 (Tension)

No directionality was observed in the response pattern. All the original items were included in the first run and no further analysis was done with the additional items.

DISCUSSION

The response patterns were similar for many of the factors, with approximately half to two thirds of the items being answered in a 'positive' direction, while the rest of the values were average or below average. In the case of Factors B and Q2, a strong 'negative' tendency was noted, while no definite direction was observed for Factors I, M, N, and Q4 with means ranging evenly from below to above average. Responses on items of all the factors varied across a broad range, thus showing potential for discrimination between respondents in terms of the different traits. Because this was a relatively small sample consisting of a specific subgroup, one cannot conclude that these response patterns are unique to the language and cultural group.

Comparisons with other cultural groups are also required, but such analyses were not feasible at this stage. However, when considering the mean raw scores for the primary factors, similar patterns are noted in the case of most of the factors for a South African

sample and also in the case of many of the factors for the American standardisation sample (Conn & Rieke, 1994; Maree, 2002; Prinsloo, 1998). Factor B is an exception where the opposite pattern was found for the latter samples.

English language proficiency was found to play a role in the results obtained on the 16PF when testing non-native speakers of that language (Abrahams & Mauer, 1999b; Maree, 2002; Prinsloo, 1998; Van Eeden et al., 1996). In an exploratory substantive analysis of the items of the 15FQ+, Meiring et al. (2005) were able to more clearly define the role of language when testing cross-culturally. The level and understanding of the words being used, the understanding of the context and interrelationships of words, the understanding of phrases and idiomatic expressions, double meanings, and qualifying words could all have affected performance. The findings of the present study show that a literal translation does not provide an adequate solution to these and related issues.

In the case of a number of the ineffective items, translation changed the meaning of the item. Reasons for this include the lack of the appropriate concept required for equivalence in the target language (e.g., 'depression' and 'upset' in the case of Factor C) and difficulty in translating idiomatic expressions (e.g., 'happy-go-lucky' in the case of Factor F). This could have resulted in misinterpretation by the respondents. Potential confusion due to the use of the negative form was noted in a number of instances (although not in all cases where this form was used) and a few translation errors occurred. Grammatical differences between the original and target languages would not have been easy to detect but could also have influenced the understanding of the items. Abrahams and Mauer (1999b) give an example of confusion that could partly have been due to this problem. An item that contains the phrase 'gets upset by events more quickly than do other people' was understood by some of the respondents in their study as a choice between events or people, as to what upsets them. Change in language usage over time should also be considered, as this could affect understanding in either the English-language or translated version of a test (Wallis & Birt, 2003).

Difficulty in the translation of personality questionnaires into an African language was also reported elsewhere. Horn (2000) refers to the restricted vocabulary in terms of personality descriptives of isiXhosa. Piedmont, Bain, McCrae, and Costa (2002) found that translating the NEO-PI-R into Xitsonga was difficult because of the absence of some of the constructs in that culture. Taylor (personal communication, N. Taylor, September 2006) refers to an attempt to translate the 16PF5 into an African language that has been temporarily abandoned due to the difficulties encountered.

Translation could result in inaccuracies and changes in the meaning of items (as illustrated in the present study), and make the items difficult to understand. However, it is clear that the problems experienced at an item level when testing in a multicultural context are not merely related to language but also involve a cultural component. Language gives expression to the concepts that are measured, and the item content is thus linked to the culture in which the test was developed (McCrae, 2000). It is therefore probable that the

content of items is culture bound, regardless of the language in which the questionnaire is presented.

Words such as ‘depression’, ‘upset’, ‘personal meaning’, and ‘motives’ are abstract concepts mostly referring to an affective component. Furthermore, if one considers those factors where trends in terms of problematic items were noticeable, it seems that in the case of Factor A (Warmth), the items dealing with feelings were mostly problematic. It is hypothesised that the construct defined as warmth probably does not manifest in this cultural group as openness in terms of the expression of feelings. This type of item might therefore not be suitable for measuring this construct with Tshivenda-speaking people. One explanation could be that, although this construct is relevant, its manifestation would be affected by the emotional reserve shown by members of the particular cultural group. Some support for this explanation was found in the case of Factor N (Privateness), where a tendency to share in close relationships was noted, while problematic items primarily dealt with being open with others in general. Possible cultural differences related to the experience and expression of emotions were furthermore noted in the relatively large number of problematic items identified in the case of Factor C (Emotional Stability) (although a specific category of items could not be identified for this factor).

Other trends, not related to an affective component, were also observed in the present study. Acculturation might have played a role in the case of the job-related items excluded for Factor A (Warmth). This could have been due to insufficient experience with these jobs. Educational background possibly played a role in the case of Factor B (Reasoning) where the respondents seemed to find the items difficult. This corresponds with findings by, among others, Van Eeden and Prinsloo (1997). In the case of Factor E (Dominance), an overlap between the constructs of affiliation and assertiveness seemed possible, with items measuring the former being excluded. This group showed a strong tendency towards group affiliation in terms of their performance on Factor Q2 (Self-reliance), and items dealing with cooperation (or the lack of it) with others are probably not suitable for measuring a construct that reflects a degree of self-centredness.

Van der Colff (2003) refers to the African value system of ubuntu as a collective mindset that embodies value sharing and communal enterprise. Blunt and Jones (1997) regard the search for harmony that is implied by this concept as the driving force in the African context. Given this tendency to interdependence, Meiring, Van de Vijver, De Bruin, et al. (2006) assume expressions of sociability and relatedness to be under strict normative control among black South Africans. They speculate that the relationship between sociability and dominance that was identified by them may therefore be less salient in these groups. This supports the hypothesis related to Factor E in the present study.

Considering the literature, it seems prudent to relate the trends found in this study to what Abrahams and Mauer (1999b) refer to as cultural norms and experiential factors — that is, differences in the manifestation of constructs rather than culturally specific constructs.

CONCLUSIONS

Based on these findings, validity analyses of the Tshivenda translation of the 16PF5 and equivalence studies were at this stage not feasible. With a few exceptions, the reliability estimates found for the Tshivenda translation were low even after the exclusion of problematic items. Furthermore, many of the items of this translation have to be changed or replaced. It is necessary to address the problems in the translation process that were identified, but it was also indicated that item content needs to be evaluated in terms of cultural appropriateness. This implies that either the adaptation or the assembly option suggested by Van de Vijver and Tanzer (1997) needs to be considered.

The first option will involve retaining many of the items, but also changing items and developing new items. Some of the additional Form S items might be used as substitutes. According to Ellis (1995), a substitute item should assess the same construct, have the same item parameters, and provide the same information at the same point on the trait continuum. A committee approach seems indicated, with experts from different cultural groups being involved. This group should at least include psychologists and language specialists. The findings of the present study could inform the process. Direct comparison of the different language versions will no longer be possible, and the equivalence of such an adapted translation to the English version of the 16PF5 will have to be established. This option might place limitations on cross-cultural research and comparability of results in practice.

The second option requires consideration of the meaning of the constructs measured by the 16PF5 in the Venda culture to ensure that items adequately represent relevant constructs. The cultural trends identified here, although valuable, will not be sufficient as a basis for this adaptation. Abrahams and Mauer (1999b) refer to the difficulty in identifying cultural trends and the potential contribution of psychologists, sociologists, and anthropologists. As with the previous option, equivalence studies are indicated. The aim will be not only to demonstrate the degree of cross-cultural equivalence, but also to interpret observed differences (Van de Vijver & Tanzer, 1997).

The ideal is a personality test that provides for different cultures at a conceptual level. Culture-specific dimensions could be included in existing tests, although the development of an indigenous personality inventory might be more appropriate for an understanding and assessment of personality in the South African context (Meiring, Van de Vijver, De Bruin et al., 2006). Such a project has been initiated with the aim of developing a South African personality inventory that provides for universal and unique dimensions of personality. However, finding a solution to cross-cultural assessment will continue to be complicated by the fact that acculturation and assimilation into Western culture influence the extent to which individuals identify with the cultures of particular ethnic groups (Marsella et al., 2000).

ACKNOWLEDGEMENTS

This article is based on the second author's master's dissertation completed at the University

of South Africa. We thank the Human Sciences Research Council and in particular Cas Prinsloo for support with this study. We also thank Cas Coetzee from the University of South Africa for his assistance with the statistical analysis. Based on the reviewer reports we were able to further develop the theoretical foundation of this study and to use this in the interpretation of the results.

REFERENCES

- Abrahams, F., & Mauer, K. F. (1999a). The comparability of the constructs of the 16PF in the South African context. *Journal of Industrial Psychology, 25*(1), 53–59.
- Abrahams, F., & Mauer, K. F. (1999b). Qualitative and statistical impacts of home language on responses to the items of the Sixteen Personality Factor Questionnaire (16PF) in South Africa. *South African Journal of Psychology, 29*, 76–86.
- Blunt, P., & Jones, M. L. (1997). Exploring the limits of Western leadership theory in East Asia and Africa. *Personnel Review, 26*(1/2), 6–23.
- Bracken, B. A., & Barona, A. (1991). State of the art procedures for translating, validating and using psychoeducational tests in cross-cultural assessment. *School Psychology International, 12*(1–2), 119–132.
- 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., & Hugo, H. L. E. (1993). *The relevance of the general Scholastic Aptitude Test (GSAT) for pupils who do not have English as their mother tongue* (Report ED-21). Pretoria: Human Sciences Research Council.
- Claassen, N. C. W., Krynauw, A. H., Holtzhausen, H., & wa ga Mathe, M. (2001). *A standardisation of the WAIS-III for English-speaking South Africans*. Pretoria: Human Sciences Research Council.
- Claassen, N. C. W., & Schepers, J. M. (1990). Groepverskille in akademiese intelligensie verklaar op grond van verskille in sosio-ekonomiese status (Group differences in academic intelligence that can be explained by differences in socio-economic status). *South African Journal of Psychology, 20*, 294–302.
- Conn, S. R., & Rieke, M. L. (1994). *The 16PF Fifth Edition Technical Manual*. Champaign, IL: Institute for Personality and Ability Testing.
- De Bruin, G. P. (2005, August). Cross-cultural psychological assessment in Africa: Lessons learnt in South Africa. Paper presented at the 8th European Conference of Psychological Assessment, Budapest.
- De Bruin, G. P., Schepers, J. M., & Taylor, N. (2005, June). Construct validity of the 16PF 5th Edition and the Basic Traits Inventory. Paper presented at the 8th Annual Industrial Psychology Conference, Pretoria.
- Ellis, B. B. (1995, August). Examination of the measurement equivalence of a Spanish version of the 16PF using item response theory. Paper presented at the 103rd Annual Convention of the APA, New York.

- Foxcroft, C. D. (1997). Psychological testing in South Africa: Perspectives regarding ethical and fair practices. *European Journal of Psychological Assessment, 13*, 229–235.
- Grieve, K. (2005). Factors affecting assessment results. In C. Foxcroft & G. Roodt (Eds.), *An introduction to psychological assessment in the South African context* (2nd ed.) (pp. 224–241). Cape Town: Oxford University Press.
- Grieve, K., & van Eeden, R. (1997). The use of the Wechsler Intelligence Scales for adults and the implications for research and practical application in South Africa. Paper presented at a national symposium on intelligence testing in South Africa, Pretoria.
- Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment, 10*(3), 229–244.
- Hambleton, R. K., & Kanjee, A. (1995). Increasing the validity of cross-cultural assessments: Use of improved methods for test adaptations. *European Journal of Psychological Assessment, 11*(3), 147–157.
- Heuchert, J. W. P., Parker, W. D., Stumpf, H., & Myburgh, C. P. H. (2000). The five-factor model of personality in South African college students. *American Behavioral Scientist, 44*(1), 112–125.
- Horn, B. S. (2000). A Xhosa translation of the revised NEO Personality Inventory. Unpublished master's dissertation, University of Port Elizabeth, South Africa.
- Kanje, A. (2005). Cross-cultural test adaptation and translation. In C. Foxcroft & G. Roodt (Eds.), *An introduction to psychological assessment in the South African context* (2nd ed.) (pp. 57–68). Cape Town: Oxford University Press.
- Laher, S. (2007). Personality reconceptualised: A Millonian approach? *South African Journal of Psychology, 37*(1), 82–95.
- Maree, D. J. F. (2002). *Properties of and South African norms for the 16PF Fifth Edition* (Technical Report). Johannesburg: Jopie van Rooyen and Partners.
- Marsella, A. J., Dubanoski, J., Hamada, W. C., & Morse, H. (2000). The measurement of personality across cultures: Historical, conceptual, and methodological issues and considerations. *American Behavioral Scientist, 44*(1), 41–62.
- McCrae, R. R. (2000). Trait psychology and the revival of personality and culture studies. *American Behavioral Scientist, 44*(1), 10–31.
- Meiring, D., van de Vijver, F. J. R., de Bruin, G.P., & Rothmann, S. (2006). Personality structure in South Africa: Commonalities of three comprehensive measures (Draft article).
- Meiring, D., van de Vijver, F. J. R., & Rothmann, S. (2006). Bias in an adapted version of the 15FQ+ in South Africa. *South African Journal of Psychology, 36*(2), 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 measures in South Africa. *South African Journal of Industrial Psychology, 31*, 1–8.
- 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*(2), 100–109.

- Nell, V. (1997). Science and politics meet at last: The insurance industry and neuropsychological norms. *South African Journal of Psychology*, 27(1), 43–49.
- Owen, K. (1991). Test bias: The validity of the Junior Aptitude Tests (JAT) for various population groups in South Africa regarding 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 Differences*, 13, 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* (Revised edition), (pp. 77–96). Pretoria: Human Sciences Research Council.
- Piedmont, R. L., Bain, E., McCrae, R. R., & Costa, P. T. (2002). The applicability of the five-factor model in a sub-Saharan culture: The NEO-PI-R in Shona. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 155–173). New York: Kluwer Academic/Plenum.
- Prinsloo, C. H. (1992). *Manual for the use of the Sixteen Personality Factor Questionnaire: South African 1992 version (16PF, SA92)*. Pretoria: Human Sciences Research Council.
- Prinsloo, C. H. (1998). *The factor structure and readability of the South African English version of the Fifth Edition of the American 16PF (Report to the External Advisory Committee)*. Pretoria: Human Sciences Research Council.
- 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(3), 48–57.
- Ralston, D. A. (1995). Cultural accommodation: The effect of language on the response of bilingual Hong Kong Chinese managers. *Journal of Cross-cultural Psychology*, 26, 714–727.
- Retief, A. (1992). The cross-cultural utility of the SAPQ: Bias or fruitful differences? *South African Journal of Psychology*, 22, 202–207.
- Russell, M. T., & Karol, D. L. (1994). *The 16PF Fifth Edition administrator's manual*. Champaign, IL: Institute for Personality and Ability Testing.
- Shuttleworth-Edwards, A. B., Kemp, R. D., Rust, A. L., Muirhead, J., Hartman, N., & Radloff, S. (2004). Cross-cultural effects on IQ test performance: A review and preliminary normative indications on WAIS-III test performance. *Journal of Clinical and Experimental Neuropsychology*, 26(7), 903–920.
- Stumpf, H. (1993). The factor structure of the Personality Research Form: A cross-national evaluation. *Journal of Personality*, 61, 27–48.
- Taylor, I. (2000). The construct comparability of the NEO-PI-R Questionnaire for Black and White employees. Unpublished doctoral thesis, University of the Free State, Bloemfontein, South Africa.
- 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.
- 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. (1996). Intelligence tests. In K. Owen & J. J. Taljaard (Eds.), *Handbook for the use of psychological and scholastic tests of the HSRC* (Revised edition), (pp. 157–190). Pretoria: Human Sciences Research Council.
- van de Vijver, F. J. R., & Poortinga, Y. H. (1992). Testing in culturally heterogeneous populations: When are cultural loadings undesirable? *European Journal of Psychological Assessment*, 8(1), 17–24.
- van de Vijver, F. J. R., & Tanzer, N. H. (1997). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, 47(4), 263–279.
- van der Colff, L. (2003). Leadership lessons from the African tree. *Management Decision*, 41(3), 257–261.
- van Ede, D. M. (1996). How to adapt a measuring instrument for use with various culture groups: A practical step-by-step introduction. *South African Journal of Higher Education*, 10(2), 153–160.
- 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* (Client 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.
- Wallis, T., & Birt, M. (2003). A comparison of native and non-native English-speaking groups' understanding of the vocabulary contained within the 16PF (SA92). *South African Journal of Psychology*, 33(3), 182–190.
- Wilson, D., Doolabh, A., Cooney, J., Khalpey, M., & Siddiqui, S. (1990). A cross-cultural validation of the Personality Research Form in Zimbabwe. *International Journal of Psychology*, 25, 1–12.

Copyright of South African Journal of Psychology is the property of South African Journal of Psychology and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.