

CHAPTER 4

CROSS-CULTURAL ASSESSMENT

4.1 Chapter Preview

Frequently cross-cultural research involves the application of instruments in various linguistic groups (Van de Vijver & Tanzer, 1997) and, in this study, an existing postpartum depression screening measure developed for the American culture, was adapted into Afrikaans. Therefore, certain issues regarding the compilation of instruments for use in cross-cultural research and cross-cultural application of tests needs to be addressed. This chapter will look specifically at cross-cultural assessment, factors influencing cross-cultural assessment, methodological considerations for cross-cultural assessment, ethical guidelines for adaptation of cross-cultural assessment measures, and translating assessment measures.

4.2 Cross-Cultural Assessment

Cross-cultural assessment is the evaluation of behaviour and attributes by obtaining measures of these under different cultural conditions and by comparing them in order to establish cross-cultural uniformities and differences (Van Ede, 1996). Irrespective of how large or small the cultural difference, cultural groups often share “a large part of their everyday life-worlds, a country, and also a common humanity” (Retief, 1988, p. 183). Knowledge of these uniformities can be used to develop a pan-human theory of human

behaviour, while knowledge of differences makes us aware of variations caused by the influence of different cultural conditions (Van Ede, 1996).

Comparative studies across ethnic groups and cultures attempt to elucidate discrepancies among human beings and thus try to achieve a better understanding of human society and behaviour. Researchers in this field label their research as transcultural, cross-national, cross-cultural, or cross ethnic. A dilemma for these comparative studies is to compile instruments that do not discriminate against individuals. Certain individuals may not have been exposed, in their ethnic, cultural or subcultural group, to the issues required by the instrument. For instance, the Minnesota Multiphasic Personality Inventory contains various implicit references to the American culture and extensive adaptations would be required before it could be used in other languages and cultures (Lucio, Reyes-Lagunes, & Scott, 1994).

The comparison of people from different cultural groups has long since become an important part of behavioural science (Manaster & Havighurst, 1972). Baron and Byrne (1994) agree that efforts to understand social behaviour must take careful account of cultural factors. Attention to the effects of cultural factors is an increasingly important trend in modern social psychology. According to Anastasi and Urbina (1997) the problem associated with assessing people who have highly dissimilar cultural backgrounds was recognized in the United States as early as 1910 when large groups of immigrants had to be assessed. The issue of cross-cultural assessment has received increasing attention since the middle of the 20th century when assessment measures were needed in newly developing nations in Africa and elsewhere to decide on admission to educational facilities and for individual counselling.

4.2.1 Multicultural assessment in South Africa.

4.2.1.1 Instrument development versus translation and adaptation.

There is considerable evidence (e.g., Van Ede, 1996; Van Eeden & Prinsloo, 1997; Van de Vijver, 2002; Van de Vijver & Poortinga, 1997; Van de Vijver & Lonner, 1995) to suggest that interest in international comparative studies of cross-cultural research is growing. With this growth has come the need to adapt (or translate) psychological instruments for use in multiple cultures and languages. This is especially pertinent in a linguistically and culturally diverse country like South Africa.

Most measures available in South Africa were developed in the United States of America or the United Kingdom and tend to be more appropriate for westernized English-speaking people (Foxcroft et al., 2006). It would seem impractical and virtually impossible to develop one measuring instrument which would be appropriate for the entire South African population. There would also be many obstacles to overcome in developing a measure suitable for all South Africans, such as

- The measure would need to exhibit appropriate levels of semantic and conceptual equivalence across cultures and languages;
- The procedures through which it is administered must minimize any problems created by lack of normative equivalence;
- The use of a multicultural team approach is likely to be extremely costly and time consuming;

- It becomes virtually impossible to make cross-national comparisons unless the instrument is translated and adapted for all South Africa's population groups in order to make cross-national comparisons.

It seems more appropriate to maximise the use of available, internationally relevant measures as far as is possible across cultural groups, rather than to embark on a totally new screening measure. Selecting an internationally well-researched measure and adapting and translating it for local conditions is also more time and cost effective. Existing measures have the advantage of being accompanied by the attributes of familiarity, experience, and often a vast body of research data. Existing measures can serve as a baseline for modification of culturally loaded test items in the South African context, and the gradual development of localized norms. Despite the advantages of translating and adapting an existing measure, there are numerous methodological issues that need to be addressed, such as bias and equivalence, and whether the measure is culture fair.

4.2.1.2 Progression of psychological assessment in South Africa.

Psychological assessment in South Africa has followed international trends. Measures were imported from overseas from the early 1900's (Foxcroft as cited in Van de Vijver & Rothmann, 2004, p.2). Claassen points out that, initially, psychological measures were developed separately for the English and the Afrikaans-speaking

populations (as cited in Van de Vijver & Rothmann, 2004, p. 2) and were only initiated with the White population (Huysamen, as cited in Van de Vijver & Rothmann, 2004, p. 2) – who were, and still are, a minority of the population group. Abrahams and Mauer argue that this discrimination meant that all population groups in South Africa were not adequately represented in the standardisation samples used to derive norm tables, and that the constructs being measured were different from those which the tests had been designed and standardised for (as cited in Van de Vijver & Rothmann, 2004, p. 2).

Biesheuvel explored the effects of potential bias problems associated with cross-cultural assessment in a South African context. He underlined the importance of schooling, home environment, and nutrition, as well as other factors on the cognitive performance on tests in a multicultural society (as cited in Van de Vijver & Rothmann, 2004, p. 2). The apartheid policy in South Africa resulted in a paucity of research on the bias and equivalence of assessment measures between 1960 and 1984 (Claassen and Owen as cited in Van de Vijver & Rothmann, 2004, p. 2). This changed in the 1980's, however, with a renewed interest in the comparison of cultural groups on various assessment measures in order to address issues of bias and equivalence. Since then concern has been expressed about the effectiveness and relevance of some assessment measures used in South Africa (Sibaya, Hlongwane, & Makunga, as cited in Van de Vijver & Rothmann, 2004, p. 2).

The first democratic elections in 1994 resulted in South Africa being regulated by a new constitution in which quality of individuals and basic human rights are guaranteed. This has also impacted on psychological assessment in South Africa and placed the cultural appropriateness of psychological tests and their usage in the spotlight. This led to

South Africa's new Employment Equity Act 55 of 1998, Section 8 (Government Gazette, 1998, as cited in Van de Vijver & Rothmann, 2004) which stipulates that

Psychological testing and other similar assessments are prohibited unless the test or assessment being used – (a) has been scientifically shown to be valid and reliable, (b) can be applied fairly to all employees, and (c) is not biased against any employee or group. (p. 1)

The expectations and demands raised by this Act puts a great deal of pressure on psychologists to ensure that tests are fair and unbiased. This would be quite a feat in a country which is as linguistically and culturally diverse as South Africa is. A primary goal for assessment professionals in South Africa is, and ought to be, to bring current practice in line with legal demands. This requires the development of new instruments and the validation of existing instruments for use in multicultural groups. This Act may ultimately “enhance the professional level of psychological practice by putting multicultural assessment on the agenda of the profession and by stimulating the development of new tests and even new testing practices” (Van de Vijver & Rothmann, 2004, p. 1).

Research in South Africa which addresses bias and equivalence of assessment measures has become an increasingly explored topic. Van de Vijver and Rothmann (2004) state, however, that “much more research is needed on the equivalence and bias of assessment tools used in South Africa before psychology as a profession can live up to the demands implied in the Equity Act” (p. 2).

4.3 Culture-Fair Tests

In order to address multiculturalism, attempts were initially made to develop tests that were culture-free (Cattell as cited in Van de Vijver, 2002, p. 546; Foxcroft et al., 2006). Classic culture-free tests were developed to eliminate the influence of parameters such as reading, speed, and language. Previously, researchers believed that measures could be developed which were free from cultural influences and could be applied in all cultures and reflect comparable findings. Anastasi and Urbina (1997) and other writers (e.g., Manaster & Havighurst, 1972) maintain that it is useless to try to devise a test that will not be affected by cultural influences. Anastasi (1988) states

We now recognize that hereditary and environmental factors interact at all stages in the organism's development and that their effects are inextricable intertwined in the resulting behavior. For man, culture permeates nearly all environmental contacts. Since all behavior is thus affected by the cultural milieu in which the individual is reared and since psychological tests are but samples of behavior, cultural influences will and should be reflected in test performance. It is therefore futile to devise a test that is free from cultural influences. (p. 345)

Researchers soon realized that it was impossible to develop a test that was free from any cultural influences and existing cultural measures should not be seen as interchangeable but rather as assisting in providing different types of cross-cultural comparisons (Foxcroft et al., 2006; Grieve, 2006; Mushquash & Bova, 2007; Plank,

2001). Tseng has expressed concern regarding how appropriate and useful it is to apply a conventional assessment instrument to individuals from diverse cultural backgrounds (as cited in Mushquash & Bova, 2007, p. 57). Further concerns are expressed by Butcher, Nezami, and Exner that, regardless of diverse cultural backgrounds, crucial decisions, and treatment plans are formulated according to the outcomes of clinical assessment tools that were developed for the general population (as cited in Mushquash & Bova, 2007, p. 57). Consequently, test developers focused more on culture-reduced, culture-fair, or culture-common tests in which the aim was to remove as much cultural bias as possible and include only behaviour that was common across cultures (Jenson and Cattell & Cattell as cited in Van de Vijver, 2002, p. 546; Foxcroft et al., 2006; Hogan, 2007).

If an instrument is translated from English, any comparison of groups rests on the assumption that test adaptation was culture-fair (Zeidner, Matthews, & Roberts, 2004). “A culture-fair test is equally appropriate for members of all cultures and comprises items that are equally fair to everyone” (Kitayama & Cohen, 2007, p. 561), in other words a culture-fair test tries to eliminate any social or cultural advantages, or disadvantages, that a person may have due to their upbringing.

According to Manaster and Havighurst (1972), a culture-fair test should have the following characteristics:

- It taps aspects of experience that are common to all people to whom the test will be administered, based on factors such as common family systems, language, objects in every day life, and number systems.

- It is designed to provoke an equal degree of intrinsic interest in subjects from the different cultural groups to whom it will be administered.
- It uses a language that is widely familiar and directions are stated in simple operational terms that are easily understood and have the same meaning for all the subjects to whom the test is administered.

It is highly improbable that any single measure could be designed that would incorporate all these characteristics if it had to be administered cross-culturally or cross ethnically. Culture-fair testing is a contentious issue and some authors believe that culture-fair testing is a myth that perpetuates xenophobic and racist agendas and they contend that adjusting for culture is not simply a matter of new norms or adjusting the interpretation of test scores. They believe that it involves an entire new set of testing skills to understand how the person views the experience and whether they understand what is expected. In addition, the researcher needs to interpret the results in the light of this understanding (Barrett & George, 2004).

Two major problems with culture-fair instruments were described by Anastasi (1988), namely, a lack of sufficient knowledge of the cultures concerned by outside designers of culture-fair instruments, and secondly, comparability becoming a matter of intuitive judgment rather than objective standardization.

For the purpose of cross cultural assessment, instruments may be grouped into three general categories (Van de Vijver, 2002):

1. Instruments with a known reliability and validity in Western groups. To what extent these measures retain their psychometric properties after translation would need to be determined empirically.
2. The development of new instruments that are designed to function in a cross-cultural context. These have been referred to as “culture-free”, “culture-fair”, and more recently, “culture-reduced”.
3. Culture-specific instruments that are developed because existing instruments are considered invalid, unreliable, and do not explore the target construct in other cultural groups. The instrument may be newly developed or based on major or minor adaptations of existing measures.

4.4 Factors Influencing Cross-Cultural Assessment

Shuttleworth-Jordan (1996) advocates that a clear distinction should be drawn between the following factors in the consideration of cross-cultural test influences:

- Racial differences (i.e., ethnic factors); and
- Socio-cultural differences (i.e., factors such as primary language, current language usage, socioeconomic status, preschool socialization experiences, levels of education, and test sophistication) as these are frequently associated with racial differences, and are known to account for significant variations in test performance.

She further stresses the importance of recognizing the complex and evolutionary nature of socio-cultural influences in planning appropriate test procedures. Ardila (1995) also points out that it is important to distinguish between the variable of formal education and the variable of culture which includes factors such as familial socialization, primary language, and meaning ascribed to tests. He further states that language is a variable which needs to be considered as a factor that can have test effects in its own right as it is strongly associated with both cultural background and level of education. Grieve (2006) states that “a test score has no meaning unless it is viewed in context” (p. 229). Hambleton (1994) agrees

There are many factors which affect cross-cultural/language comparisons which need to be considered whenever two or more groups from different language/cultural backgrounds are compared, especially when an instrument is being developed or adapted, or scores are being interpreted. However, often it is necessary that some of these factors are *not* merely taken into account, but that practical steps be taken to either minimize or eliminate the likely (unwanted) effects of these factors on any cross-cultural/language comparisons that are made. (p. 233-234)

Hambleton (1994) identified some pertinent factors in the social context which include schooling, language, culture, and environmental factors. This section will review some of the factors affecting cross cultural assessment.

4.4.1 Schooling.

The level of schooling attained and, in the South African context, the quality of education received indirectly influences the outcome on intelligence measures (Grieve, 2006; Nell, 2000). Holding et al. (2004) found that there is a strong relation between scores on intelligence measures and scholastic and academic achievement. However, in South Africa this situation is further complicated due to the apartheid regime, where the previously disadvantaged learners received a poorer quality of education than their privileged counterparts (Grieve, 2006; Shuttleworth-Edwards et al., 2004).

4.4.2 Language.

Language is regarded as the most important moderator of performance on assessment measures (Grieve, 2006; Nell, 2000). Poor performance on a measure could be attributed to language difficulties as opposed to ability if the measure was administered in a language other than the test-takers home language. Generally even if you are bilingual, it takes longer to process information in another language. Furthermore, according to the American Educational and Research Association (AERA), the APA, and the National Council on Measurement Education (NCME) an individual who knows two languages may not test well in either of them (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). We can think and discuss so much better in our own language and hence would more than likely perform better in a test that is written and

administered in our home language than in a test in a second language. Thus, language becomes a potential source of bias.

Translating the measure could offer a solution. This, however, could also pose difficulties such as some languages do not have the concepts and expression required by a measure. Further, translating items could affect their level of difficulty. Another complication in South Africa is that some learners are schooled in a language other than their home language, which may compromise both languages, and places test-takers in a doubled disadvantage situation (Grieve, 2006).

Respondents need to understand the language of the assessment measure and need to respond by means of language, as most measures require the use of language. Respondents who are not proficient in the language of the test may introduce construct irrelevant components to the testing process (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999).

Cultural groups may differ in their language spoken. They may also differ in terms of the way in which verbal expressions are formally structured, even if they speak the same language. Some cultural groups, for example upper-middle class North Americans and North Europeans, encourage a highly structured, rational, and orderly use of language, while other cultural groups use language more loosely, with less logical structure and less clear-cut meaning. Furthermore, different cultural groups may assign different meanings to commonly used expressions. Respondents from one cultural or ethnic group will therefore differ to other cultural or ethnic groups in their performance to

the extent that they are familiar with the questionnaire's language as well as expressions associated with that language.

Hay (2002, p. 23) considered the cultural diversity of her South African sample and points out that

it is sweeping and broad to assume cultural differences only between Black and White South Africans. There are probably as many cultural differences between English and Afrikaans Westerners, between English-speakers and speakers of other European languages, between Zulus and Sothos, and so on. It is a mistake to assume that little cultural difference exists where groups speak the same language and believe in the same God.

4.4.3 Culture.

Culture may be defined as the learned attitudes and behaviour that are characteristic of a particular social group or organization which is passed from generation to generation (Ponterotto, Casas, Suzuki, & Alexander, 1995). Our culture influences the way we learn, think, and behave. It is an integral part of our environment and cannot be isolated as a factor on its own. Further, the content of any measure reflects the culture of the test developer and the country in which it is to be used. Therefore test-takers who do not share this culture will be at a disadvantage. As discussed in section 4.2, there are no culture-free measures but practitioners are expected to be sensitive to cultural fairness in assessment and not assume equivalence between cultures. A further problem is that there

are variations in acculturation, which refers to the “process by which people become assimilated into a culture” (Grieve, 2006, p. 232).

Researchers have for instance, found cultural differences with respect to child birth, neonatal care, and infant and child rearing practices (Anastasi, 1988; Rebelsky & Daniel, 1976). These differences between various cultural and language groups are a function of not only the different traditions, norms, and values, but of different worldviews and interpretations as well (Hambleton, 1994). It is therefore entirely possible that the same construct is interpreted and understood in completely different ways by two different groups. The concept of intelligence, for example, is known to exist in almost all cultures. Lonner found that in many Western cultures this concept is associated with being able to produce responses very quickly, whereas in Eastern cultures, intelligence is often associated with slow thoughtfulness, reflection, and saying the right thing (as cited in Van de Vijver & Poortinga, 2005, p.39). Researchers in the past have neglected to consider these differences and made fallacious assumptions about individuals belonging to different cultures (Bhamjee, 1991).

4.4.4 Environmental factors.

Environmental factors determine the types of learning experiences and opportunities to which we are exposed, which in turn, affects our ability and the level to which we use that ability. Environmental factors can be grouped into distal factors (e.g., socio-economic status and enriching social environment) and proximal factors (e.g., socialization experiences in the home; Grieve, 2006).

4.4.4.1 *The home environment.*

Certain child rearing practices have been linked to promoting development of competence and cognitive abilities. These include parental responsiveness and the provision of home stimulation (Grieve, 2006).

4.4.4.2 *Socio-economic status.*

Socio-economic status refers to the person's social standing. The major indicators of SES are education, occupation, and income. The test-takers SES is important as it determines the type of facilities that are available (e.g., schools, libraries, clinics, and other social services), the opportunities that present themselves, and the attitudes of others (Grieve, 2006).

4.4.4.3 *Urbanization.*

Urbanization is generally found to influence cognitive scores, with urban children outperforming their rural counterparts (Mwamwenda, 1995). The reasons for this could be attributed to an invigorating urban environment that stimulates cognition, access to education at an early age, higher parental levels of educations, and so forth (Grieve, 2006).

4.5 Methodological Considerations in Cross-Cultural Assessment

Both bias and equivalence are fundamental concepts in cross-cultural assessment, in that they refer to the characteristics of a cross-cultural comparison of an instrument rather than the intrinsic properties. Bias and equivalence are concepts that are closely related. Bias refers to “factors that show a differential impact on scores in cultural populations, while equivalence involves the implications of bias on the scope for comparing scores” (Van de Vijver, 2002, p. 548). The equivalence of a measure (or lack of bias) is a prerequisite for valid comparisons across cultural populations (Van de Vijver & Tanzer, 1997), if bias occurs, the equivalence of the scores is challenged (Van de Vijver, 2002). Throughout the history of psychological research there have been many sweeping generalizations about differences in traits and abilities of cultural populations. When examined more closely, however, these generalizations were often based on inadequate psychometric measures. To avoid such blundering statements it would be advisable to demonstrate the absence of bias (i.e., equivalence) instead of simply making the assumption (Poortinga & Malpass, 1986). It is imperative to determine the cultural appropriateness of an instrument and there should be an empirical investigation into the item bias, differential item functioning, and construct equivalence for the different subgroups (Foxcroft et al., 2006).

4.5.1 Bias.

Bias or specifically, test bias, refers to “whether a measure is differentially valid for different subgroups” (Foxcroft et al., 2006, p. 5). Bias occurs when score differences for a

construct are observed which do not correspond to differences in the underlying trait, attitude, or ability across cultural groups (Van de Vijver, 2002; Van de Vijver & Tanzer, 1997). Bias challenges the construct validity of an item or measure (Van de Vijver, 2002). Thus, it is imperative when adapting an instrument, that any unfair advantage or disadvantage to a test-taker, irrespective of their cultural, social, economic, or linguistic background, is eliminated (Foxcroft et al., 2006). There are three types of bias: construct bias, method bias, and item bias (Van de Vijver, 2002; Van de Vijver & Leung, 1997a, 1997b; Van de Vijver & Poortinga, 1997). These types of bias will be discussed in more detail below.

4.5.1.1 Construct bias.

Construct bias occurs if the construct measured as a whole (e.g., postpartum depression) is not identical across cultural groups (Ægisdóttir et al., 2008; Van de Vijver & Tanzer, 1997). For example, the appropriateness of the item content differs between the two language versions of the measure (Ægisdóttir et al., 2008).

4.5.1.2 Method bias.

Method bias stems from the characteristics of the measure or from its administration (Ægisdóttir et al., 2008; Van de Vijver, 2001; Van de Vijver & Leung, 1997a, 1997b; Van de Vijver & Poortinga, 1997). Three types of method bias exist, namely, sample bias; instrument bias; and administration bias. Sample bias occurs when

the samples used differ in a variety of relevant characteristics other than the intended construct. Administration bias includes all sources of bias that are caused by administering the instrument (e.g., interviewee is not fluent in the language of the test). Instrument bias refers to biases that occur due to the characteristics or design, or both, of the instrument itself (Van de Vijver & Tanzer, 1997).

4.5.1.3 Item bias

Item bias or differential item functioning refers to distortions at the item level (Ægisdóttir et al., 2008). Biased items have a different psychological meaning across cultures. This has an impact on the comparison of total test scores across cultures. Thus individuals from different groups who have the same ability have a different probability of getting the item right.

Item bias may occur as a result of various factors although common causes are poor translation, poor item formulation (e.g., complex wording; ambiguity in the original item which may elicit different interpretations), low familiarity with the item content in specific cultures or inappropriateness of the item content for certain cultures, or in other words, not be equally relevant or appropriate for the cultural groups being compared (Ægisdóttir et al., 2008; Malpass & Poortinga, 1986; Van de Vijver & Poortinga, 1997; Van de Vijver & Tanzer, 2004), or the influence of some things which are considered culture-specific, for example connotations associated with the item wording or nuisance factors (Van de Vijver & Tanzer, 2004). Van Haften and Van de Vijver provide an example of item bias which was caused by inappropriate item content. The item "watched

more television than usual" had to be removed from a Western coping questionnaire when it was applied to Sahel dwellers who did not have electricity in their homes (as cited in Van de Vijver & Tanzer, 2004). Most studies of bias focus on exploring and testing for item bias (Van de Vijver & Tanzer, 1997).

Van de Vijver and Tanzer (2004) discuss strategies to identify and reduce the three types of bias mentioned above. These strategies are summarised in Table 7 below.

Table 7 Strategies for Identifying and Dealing with Bias in Cross-cultural Assessment

Type of Bias	Strategies
Construct bias	<ul style="list-style-type: none"> • Decentering (i.e., simultaneously developing the same instrument in several cultures) • Convergence approach (i.e., independent within culture development of instruments and subsequent cross-cultural administration of all instruments)
Construct bias and/or method bias	<ul style="list-style-type: none"> • Use of informants with expertise in local culture and language • Use samples of bilingual subjects • Use of local surveys (e.g., content analyses of free response questions) • Non-standard instrument administration (e.g., “thinking aloud”) • Cross-cultural comparison of nomological networks (e.g., convergent/discriminant validity studies, monotrait–multimethod studies, connotation of key phrases)
Method bias	<ul style="list-style-type: none"> • Extensive training of administrators (e.g., increasing cultural sensitivity) • Detailed manual/protocol for administration, scoring, and interpretation • Detailed instructions (e.g., with sufficient number of examples and/or exercises) • Use of subject and context variables (e.g., educational background) • Use of collateral information (e.g., test-taking behavior or test attitudes) • Assessment of response styles • Use of test–retest, training, and/or intervention studies
Item bias	<ul style="list-style-type: none"> • Judgmental methods of item bias detection (e.g., linguistic and psychological analysis) • Psychometric methods of item bias detection (e.g., differential item functioning analysis) • Error or distracter analysis • Documentation of “spare items” in the test manual which are equally good measures of the construct as actually used test items

(Van de Vijver & Tanzer, 2004, p. 128).

4.5.2 Equivalence.

The attainment of equivalent measures is perhaps the central issue in cross-cultural comparative research (Van de Vijver, 2001; Van de Vijver & Leung, 1997b). Equivalence refers to the level at which the item or test scores can be compared across cultural or language groups (Van de Vijver, 2001; Van de Vijver, 2002). For measures to be equivalent, individuals with the same or similar ability on a construct should obtain the same or similar scores on the different language version (e.g., translation equivalence) of that instrument otherwise the instrument is considered biased and the two versions of the instrument are non-equivalent. For example, a score of 10 on an unbiased scale for depression has the same psychological meaning in all cultural or language groups studied (Van de Vijver, 2002). Instruments need to be equivalent if meaningful comparisons are to be made between the two subgroups (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999; Kanjee, 2006). Without demonstrated equivalence, numerous rival hypotheses (e.g., poor translation) may account for observed cross-cultural differences (Ægisdóttir et al., 2008).

There are three types of equivalence, construct equivalence, measurement unit equivalence, and scalar equivalence (Van de Vijver & Leung, 1997a, 1997b; Van de Vijver & Poortinga, 1997). In addition, linguistic differences can easily invalidate the results of a study. If a psychometric measure is poorly translated, it doesn't matter how sound your methodology is (Onkvisit & Shaw, 2004). Therefore, it is equally important to discuss linguistic equivalence. These types of equivalence will be discussed in more detail below.

4.5.2.1 *Construct equivalence.*

Construct equivalence (also referred to as functional equivalence and structural equivalence) means that the same underlying psychological construct is measured across all cultural groups in spite of whether or not the measurement of the construct is based on identical instruments across all cultures (Van de Vijver & Tanzer, 2004).

4.5.2.2 *Measurement unit equivalence.*

Measurement unit equivalence refers to the level of equivalence that can be obtained when two metric measures have the same measurement unit but have different origins across groups (Ægisdóttir et al., 2008; Van de Vijver & Tanzer, 2004). For example, the two language version of a measure may appear the same, but equivalence is threatened if the two groups vary in their familiarity with Lickert-type answer format (method bias). Similarly, if the two groups vary in response style (acquiescence), a score of 4 on a 5-point scale may not have the same meaning for the two groups (Ægisdóttir et al., 2008).

4.5.2.3 *Scalar equivalence.*

The highest level of equivalence is scalar equivalence, which can be obtained when two metric measures have the same measurement unit and the same origin. This type of equivalence assumes completely bias-free measurement (Van de Vijver & Tanzer, 2004).

4.5.2.4 *Linguistic equivalence.*

Linguistic equivalence must be ensured when cross-cultural studies are conducted in different languages. Linguistic equivalence requires the research to pay particular attention to potential translation problems. It is therefore recommended that translators pay attention to idiomatic vocabulary, grammatical, and syntactical differences in language, as well as the experiential differences in cultures as expressed in language (Onkvisit & Shaw, 2004). The various translation techniques that can be employed to enhance equivalence will be discussed in section 4.6.

4.6 Ethical Guidelines for Adaptation of Cross-Cultural Assessment Measures

With increased globalisation and the substantial costs of test development, the use of westernized psychological tests in South Africa is widespread, therefore, having measures that are reliable and valid, and can be used for our diverse languages and cultures is crucial. However, the guidelines or standards for the translation and cultural adaptation of instruments that have been established in psychology to date are limited (Van Widenfelt, Treffers, De Beurs, Siebelink, & Koudijs, 2005).

Anastasi (1988) suggested three approaches to the development of tests for different cultures or subcultures. The basic approach is to compile an instrument that taps aspects of cultural experience which are common to many cultures, and validating the resulting measure against local criteria in the cultures where it will be administered.

Without the necessary precautions it cannot be assumed that a test is relatively free from culturally restricting elements, yet this repeated validation in various cultures has often either been neglected or poorly controlled. However, it is unlikely that a measure would fully meet these requirements across a wide range of cultures or ethnic groups.

A second approach is to compile a measure within one culture and then administering it to individuals from different cultural groups. This procedure is typically followed when the object of assessment is the prediction of a local criterion within a specific culture. The criterion itself is usually culturally loaded, therefore the test validity may drop if the cultural loading of the test is reduced. Care should be taken not to regard a measure constructed within a single culture as a universal yardstick for measurement. This approach enables the researcher to determine the cultural distance between groups as well as the individual's degree of acculturation.

The third approach is to construct different measures within each culture and to validate them against local criteria only. In this type of approach an individual's result are compared to local norms and no cross-cultural comparisons are attempted. This seems to negate the purpose of cross-cultural and cross-ethnic research.

The International Test Commission (ITC) was formally established in 1978 is an "Association of national psychological associations, test commission, publishers and other organizations committed to promoting effective testing and assessment policies and to the proper development, evaluation and uses of educational and psychological instruments." (International Test Commission, 2001). The following definition of an instrument adaptation guideline was proposed by the ITC (Hambleton, 1994)

An instrument adaptation guideline is a practice that is judged as important for conducting and evaluating the adaptation or parallel development of psychological and educational instruments for use in different populations. (p. 233)

Hambleton (1994) predicted that substantially more adaptations might be expected in the future as:

- international exchanges of tests and instruments become more common;
- credentialing exams are adapted for use in multiple languages; and
- interest in cross-cultural research increases.

In 1992 the ITC began a project to prepare guidelines for translating and adapting psychological instruments and other tests, as well as establishing score equivalence across different language or cultural groups. In 2000 the ITC released *Guidelines for Adapting Educational and Psychological Tests* (Hambleton, 2001; International Test Commission, 2001). These guidelines have become the benchmark for cross-cultural test adaptation around the world (Foxcroft et al., 2006). The ITC subsequently developed the “International Test Commission Guidelines for Translating and Adapting Tests – Version 2010” (International Test Commission, 2010) which further addresses issues of fairness and bias in test use and sets standards for the professional practice of assessment. These guidelines emphasize the importance of cross-cultural validity of measures as well as their constructs across different populations and cultures. The 22 guidelines for adapting

psychological and educational tests or instruments are organized into four categories (International Test Commission, 2010):

1. Context, which addresses concerns about construct equivalence in the language groups of interest;
2. Test development and adaptation, which includes the guidelines which arise in the process of adapting an instrument, from selecting translators to statistical methods for analyzing empirical data to investigate score equivalence;
3. Administration, which addresses guidelines having to do with the ways in which instruments are administered in multiple language groups. This includes everything from selecting administrators, to the choice of item formats, to establishing time limits; and
4. Documentation/score interpretations.

The ITC uses the term “adaptation” rather than “translation”. This is because the former term is broader and instrument adaptation guidelines seemed to more accurately reflect the process of preparing a test or instrument for use in a second language or culture. Translation is always part of the adaptation process, but is only one of a number of steps that must be carefully carried out to produce a test or instrument that is equally valid in two or more languages and cultures.

Several professional bodies have since provided clear standards and guidelines that need to be adhered to when using psychological tests, these include the Standards for Educational and psychological Testing (American Educational Research Association,

American Psychological Association, and National Council on Measurement in Education, 1999) and the Guidelines for Computer-based Tests and Interpretations (APA, 1986).

In an attempt to address issues of fairness and bias in test use, the need arose to develop standards for the professional practice of testing and assessment. Led by Bartram from the United Kingdom, the ITC developed “International Guidelines on Test Use - Version 2000” (International Test Commission, 2000), which like in many other countries was adopted by South Africa.

The international guidelines for test use (International Test Commission, 2000, p. 12) states that when individuals from diverse groups (e.g., groups differing in terms of age, gender, education, cultural background, or ethnic origin) are assessed all reasonable efforts should be made to ensure that the following best practice guidelines are followed:

- The tests are unbiased and appropriate for the various groups that will be tested.
- The constructs being assessed are meaningful in each of the groups represented.
- Evidence is available on possible group differences in performance on the test.
- Evidence relating to differential item functioning (DIF) is available, where relevant.
- There is validity evidence to support the intended use of the test in the various groups.
- Effects of group differences not relevant to the main purpose (e.g., differences in motivation to answer, or reading ability) are minimised.

- In all cases, Guidelines relating to the fair use of tests are interpreted in the context of local policy and legislation.

Ethical guidelines have also been issued (APA, 1993; International Test Commission, 2000). It has, however, been reported that practitioners, test developers, and test publishers generally do not adhere to these guidelines. Furthermore, many measures which have been translated have not been re-normed and re-validated, and there are no appropriate precautions provided in the test manuals (Barrett & George, 2004).

The ITC guidelines (International Test Commission, 2000, p18) for multilingual instruments require the following:

- Each language or dialect version has been developed using a rigorous methodology meeting the requirements of best practice.
- The developers have been sensitive to issues of content, culture, and language.
- The test administrators can communicate clearly in the language in which the test is to be administered.
- The test-taker's level of proficiency in the language in which the test will be administered is determined systematically and the appropriate language version is administered or bilingual assessment is performed, if appropriate.

The standards for educational and psychological testing (Standard 9.7; American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999) state

When a test is translated from one language to another, the methods used in establishing the adequacy of the translation should be described, and empirical and logical evidence should be provided for score reliability and the validity of the translated test's score inferences for the uses intended in the linguistic groups to be tested. (p. 99).

Standard 9.9 states that if multiple language versions are intended to be comparable, then empirical evidence of test comparability should be reported. The comprehensive set of 22 guidelines provided by the ITC for improving the translation and adaptation of educational and psychological instruments are presented in Appendix A. Hambleton (2004, pp. 65-70) summarized these guidelines and notes the following nine key steps that should be addressed when adapting or translating any assessment instrument:

1. Explore the construct equivalence in the languages and cultures of interest.
2. Decide whether test adaptation or translation is the best option.
3. Choose well qualified translators.
4. Translate or adapt the instrument using the appropriate design.
5. Review the adapted version and make the necessary changes.
6. Conduct a small pilot with the adapted test.

7. Conduct a validation investigation.
8. Place the scores of both the translated and original instruments on a common scale.
9. Document the process and prepare the manual for test users.

4.7 Translating Assessment Measures

In South Africa, many of the psychological measures are in English and translating and adapting these measures would alleviate some of the biases associated with psychological tests. This section looks at the current methods researchers employ in multilingual studies.

4.7.1 Techniques in translating instruments.

Test translation refers to the process of converting a measure from one language to one or more other languages (e.g., from English to Afrikaans), while still retaining the original meaning (Foxcroft et al., 2006). Translating a psychological instrument is more complex than simply rewriting the text into a different language (Bracken & Barona, 1991; Brislin, 1980, 1986; Geisinger, 1994; Hambleton, 1994), it needs to take into consideration the original context of the source instrument as well as reflect the cultural understanding of the target language (Bracken & Barona, 1991). Therefore, an appropriate translation requires a balance between psychological, linguistic, and cultural considerations (Hambleton, 1994; Van de Vijver & Hambleton, 1996). Employing a

proper translation methodology is critical as it affects the equivalence of the multilingual versions and the measures' cross-cultural validity. Further, researchers should also bear in mind that test instructions need to undergo the same translation method as the items (Ægisdóttir et al., 2008).

Numerous techniques have been developed for translating, adapting, and re-norming psychological instruments for cultures and languages other than their initial target population (Ferraro, 2002; Fletcher-Janzen et al., 2000; Nell, 2000). According to Bracken and Barona (1991) the successful translation of tests is dependent on following a comprehensive multistep translation and validation process. The translation techniques and processes they describe are (a) source to target language translation, (b) blind back-translation, (c) translation – back-translation repetition, (d) committee review, (e) pilot testing, (f) field-testing, and (g) norm development. Brislin (1980) proposed translation methods, such as back-translation, bilingual, committee, decentering, and pretests. Onkvisit and Shaw (2004) refer to the following translation techniques: back translation, parallel-blind translation, committee approach, random probe, and decentering. However, Van de Vijver and Tanzer (1997) refer to only two translation procedures, namely, the translation-back-translation procedure and the committee approach. Bracken and Barona (1991) advocate that the most commonly applied technique is the back-translation technique while Kanjee (2006) states that the common designs used are forward-translation and back-translation. This section will discuss some of the proposed translation techniques.

4.7.1.1 One way or bilingual translation.

This involves the translation from the original to the target language by a translator who is truly bilingual and also sufficiently educated to be familiar with the concepts of the test and subject matter (Bracken & Barona, 1991). The test translator must also be knowledgeable with the target culture, the construct being assessed, and the principles of assessment (Hambleton & de Jong, 2003; Van de Vijver & Hambleton, 1996). This will assist in minimizing item biases that may result from literal translations. This technique is akin to first step in Bracken and Barona's (1991) multistep translation process mentioned earlier. It is considered to be an uncomplicated translation method. In some instances the researcher may opt to have a few translators conduct a one-way translation of the instrument. This method is less time consuming and less expensive than other methods. Limitations of this method are, however, that no comparison of the final survey version is made and information may be lost through literal translation (McGorry, 2000).

4.7.1.2 Forward-translation.

A forward translation or simple direct translation simply implies translating the instrument into the language chosen (e.g., from English into Afrikaans), by a single translator or a group of translators (Kanjee, 2006; McGorry, 2000). Although this method is more cost effective, there may be a loss of information through literal interpretation (McGorry, 2000). This first version would then be given to a pilot group of test-takers to answer and then the test-takers would be questioned by judges as to the meaning of their responses. The judges would then decide if the responses given reflect a reasonable

representation of the test items in terms of cultural and linguistic understanding. If a high percentage of the test-takers present a reasonable representation of an item, the item is regarded as being equivalent. A valuable advantage is that the functioning of any item is provided directly by the test-takers (Kanjee, 2006). The disadvantage, however, is that there are many confounding variables (e.g., personal, cultural, linguistic) that may affect this process and hinder the results. Another disadvantage is that this technique is very labour intensive and time-consuming (Kanjee, 2006).

4.7.1.3 Modified direct translation.

Geisinger (1994) proposed that some steps be taken to overcome some concerns with the forward translation. He also suggested that a panel or committee of experts do independent checks on the original translation as follows: “(a) review the items and react in writing, (b) share their comments with one another, and (c) meet to consider the points made by each other and to reconcile any differences of opinion” (p. 306). He further recommends that the original translator meet with the panel on two occasions, first so that the panel members can explain their concerns, and then again so that the translator can give explanations and clarify why the measure was drafted as it was.

Modified direct translation is more informative than a simple forward translation due to opportunity for discussion among committee members as well as discussions between the translator and the committee. This procedure increases the security of the translation, provided the translator and committee members are competent and are able to reach consensus in the meetings. This technique is, however, likely to consume

substantial amounts of effort, time, and money. As a result, practicality is somewhat low (Behling & Law, 2000). This approach also does not involve enough collaborative efforts that are needed to produce a well-translated instrument (Pan & De La Puente, 2005).

4.7.1.4 *Parallel blind translation.*

In a parallel-blind translation, the measure is translated by several translators independently and then the translators meet to compare their versions and resolve any differences. Once the differences are resolved they jointly present the translated instrument to the researcher (Behling & Law, 2000; Onkvisit & Shaw, 2004). According to Guthery and Lowe (1992) the parallel blind technique has two advantages: speed and researcher control. The process is faster than conventional back-translation because the translators work in parallel as opposed to in sequence. However, this technique lacks source language transparency, in that unless the researcher is bilingual his or her role in the process is limited. For instance, the researcher would not be able to identify if the translators share certain misconceptions or if what appears to be an agreement between them is actually due to their unwillingness to criticize one another's translations (Behling & Law, 2000).

4.7.1.5 *Committee approach.*

A committee (or cross-translation) approach is where a group of experts (such as cultural, linguistic, and psychological) prepare a translation (Nasser, 2005; Van de Vijver

& Tanzer, 2004). The committee approach differs from the parallel-blind technique due to the fact that the former allows committee members to discuss the research questions with each other during the translation (Onkvisit & Shaw, 2004). If all the translations are the same, then the translation is considered valid (Nasser, 2005). Often researchers combine the committee approach with the back translation technique (Van de Vijver & Leung, 1997b). Major strengths in using this approach is that this collaborative effort improves the quality of the translation, reduces bias, and reduces misconceptions that a single person may bring (Ægisdóttir et al., 2008) This is especially true, if the members have complimentary areas of expertise (Van de Vijver & Tanzer, 2004). A disadvantage of this approach is the absence of an independent evaluation of the adequacy of the translation. If the researcher is not fluent in the translated language, then additional evidence will be needed to evaluate the quality of the committee's work (Van de Vijver & Leung, 1997b). In addition, translators may be reluctant to criticize one another or may miss information relevant to the intended group due to similar cultural backgrounds and education (McGorry, 2000).

4.7.1.6 Pilot-testing or pretest.

Once a translated version of an instrument has been agreed upon and approved by the bilingual review committee, it can be very beneficial to administer it to a small group of people representative of the target population for pilot testing and minor adjustment. The results can be compared to the results obtained from the original language sample (Hambleton & Patsula, 1999; Van Widenfelt et al., 2005). A trained examiner who is

fluent in the target language should administer the measure. Pilot testing is a helpful step in instrument adaptation, however, a review by Guillemín, Bombardier, and Beaton (1993) indicates that a pretest or pilot test is done by comparatively few studies on translated measures. An interview or discussion with the pilot participants is also a valuable step as it allows the examiner to determine the pilot participants' reactions to the test instructions, response categories, and items. The examiner should also take note of verbal and non verbal expressions, such as looks of puzzlement, confusion, giggles, or other responses to items that may indicate that the item is perceived as confusing, bizarre or inappropriate, and hence suggest possible translation failure (Bracken & Barona, 1991). These should be discussed by the review committee to explore possible reasons for the inappropriate examinee responses.

Pilot participants that are selected should vary in social and economic background, geographic region, gender, and age. Regional differences in expression should also be considered when using the feedback of pilot participants (Bracken & Barona, 1991; Van Widenfelt et al., 2005). After the initial pilot testing data has been obtained it is essential to meet again as a team to consider any necessary changes. Further adaptations can be made by the team of translators based on the outcome of the pilot testing (Geisinger, 1994). If uncertainty remains about an item, two alternatives for that item can temporarily be included in the version for further testing (Van Widenfelt et al., 2005).

4.7.1.7 *Field-testing.*

Field testing typically follows the pilot testing. This procedure is essentially the same as pilot testing but differs in magnitude. Examiners should be attentive of any consistent problems related to specific items or test directions. Formal item analyses can be done on the results obtained from the field testing if the sample is large enough. (Bracken & Barona, 1991).

4.7.1.8 *Random probe.*

A random probe entails placing probes at random locations in both the original and translated measures during pretesting in order to ensure that the respondents understand the items in the same way (Onkvisit & Shaw, 2004). The respondents are then asked to explain why they responded as they did to certain items (Behling & Law, 2000). According to Guthery and Lowe (1992), “if the respondent’s justification to the original answer is strange, then the intent of the question is not being conveyed” (p. 10). This technique is cheap, simple, and quick. However, researchers may need to supplement this technique with a more rigorous procedure, as it provides limited information and is not innately source language transparent (Behling & Law, 2000).

4.7.1.9 *Decentering.*

Decentering is defined by Eremenco, Cella, & Arnold (2005) as:

A process in which the source instrument and its translation are open to modification in an iterative process, so that the meaning is equivalent between them. The opposite is to have a translation process in which the source is unchangeable, thereby requiring all adjustments to be made to the translation only. (p. 228)

This translation technique is termed “decentering” because the researcher does not center in either the original language or the target language. Decentering modifies the back-translation technique by considering the original and translated versions as equally important and both are open to modifications (Beck et al., 2003; Ægisdóttir et al., 2008; Geller, Vinokurov, & Martin, 2004; Flores, 2006). If problems are identified in the original measure (e.g., words in the original language that have no equivalence in the translated language), then it should be modified in order to be more easily translatable (Ægisdóttir et al., 2008; Nasser, 2005; Onkvisit & Shaw, 2004). There is constant comparison between the two measures and the original measure is retrospectively modified in order to enhance its translatability. Thus in this process the original measure becomes a draft and is revised to fit the new research situation (McGorry, 2000). Typical modifications made are to words or concepts that are difficult to translate or are culture specific (Van de Vijver & Leung, 1997b). Marin and Marin acknowledge that the use of decentering may lengthen the translation process but it does help to achieve a fully equivalent language version (as cited in Beck et al., 2003, p. 68).

Decentering has a number of advantages. Like back-translation, it is both informative and source language transparent. Further, it provides the researcher with the

opportunity to check the reasoning of one translator against another and allows for better equivalence because the source and target versions are equally subject to change, thereby bringing both closer in meaning to the desired concept. The decentering technique, according to Van de Vijver and Leung (1997b), echoes “the goals of the culture-free and culture-fair test movement” (p. 39). This technique is not, however, very practical as it is labour intensive requiring a substantial number of translators that are multicultural, multilingual, and have expertise in the construct under study (Behling & Law, 2000; McGorry, 2000; Van de Vijver & Leung, 1997b). Furthermore, in instances where an instrument has already been validated and is widely used, the instrument developer is likely to be averse to implementing changes to the measure in light of translations. This may be the case regardless of the possibility that decentering would improve the equivalence of the translations with the established version (Eremenco et al., 2005).

4.7.1.10 Back-translation.

The back-translation method, also known as the double translation method, involves the translation of items from the original into the target language by one or more bilingual translators. This material is re-translated back into the original language by another bilingual translator or team of bilingual translators, yielding the back-translated version. Richard Brislin, a cross-cultural psychologist, was the first to write extensively about back translation as a method to ensure a quality translation of a test or measuring instrument (Brislin, 1970; 1976; 1986; Brislin, Lonner, & Thorndike, 1973; Stansfield & Bowles, 2007).

In order to judge the quality and determine the equivalence of the measures, the researcher checks for errors between the original and back-translated versions of the measure and consults with the translators about the possible reasons for any inconsistencies, mistranslations, lost words, and changes in meaning. Once these issues have been addressed further revisions to the translated version may occur. It is then back-translated yet again and compared to the source document. This process of creating a back translation and comparing it with the original version is repeated until the two versions agree. Although the researcher may go through several rounds of revisions, the original-language version of the measure is considered the standard against which the translated version is compared (Beck et al., 2003; Brislin, 1970, 1986; Geller et al., 2004; Kanjee, 2006; McGorry, 2000; Nasser, 2005; Onkvisit & Shaw, 2004). Marin and Marin (as cited in Beck et al., 2003, p. 68) point out that this is what differentiates back-translation and decentering. Decentering regards the original language version as well as the translated version as equally important.

This method has been widely applied and can identify various kinds of errors. Various researchers have stated that the back-translation method is particularly useful for checking the semantic equivalence of the translations of measures in different languages used in cross-cultural research (Beck & Gable, 2003; Bracken & Barona, 1991; Prieto, 1992). The back-translation technique has been used successfully to translate from English to Afrikaans (e.g., Shillington, 1988), as well as in developing a Spanish version of the PDSS (Beck & Gable, 2003). An advantage of using the back-translation technique is that it enhances the reliability and accuracy of the translated instrument as it offers

opportunity for revisions through the translation process (Bracken & Barona, 1991; Geisinger, 1994; Van de Vijver & Hambleton, 1996; Van de Vijver & Leung, 2001).

Some researchers have raised concerns regarding translating and adapting an instrument from one language to another (Geisinger, 1994; Hui & Triandis, 1985; Van de Vijver & Hambleton, 1996). Some problems with the back-translation method have been identified. Lack of agreement between the source document and the back translation may be due to problems with the back translation as opposed to problems with the initial forward translation. The back translation is just as likely to contain translation errors (omissions, mistranslations, insertions) as is the forward translation (Stansfield & Bowles, 2007).

When the translator is aware that the forward translation will be validated by a back translation, it may influence the translator's approach to forward translation. The translator may produce a very literal forward translation to help ensure that the back translation will produce a document that is very similar to the original version. This type of translation is, according to Stansfield and Bowles (2007, p. 2) "likely to produce stilted rather than natural expression and result in a test that is difficult to read, and hence less accessible to the examinee population."

If the back-translation version seems to lack equivalence in meaning to the original version, it is not easy to determine whether the differences are as a result of poor translation, or cultural and linguistic differences inherent in cross-cultural research. Furthermore, when the translated version is similar to the original version, it remains uncertain about the nuances of meaning across languages and cultures (Geller et al.,

2004). Back-translation may lack equivalence in meaning and still demonstrate spurious lexical equivalence, thus giving the researcher a false sense of security (Birbili, 2000; Brislin, 1970, 1976).

A translated instrument, while linguistically correct, may have poor quality from a psychological perspective. An example provided by Hambleton (1994, p. 235) that illustrates this point, is the test item “Where is a bird with webbed feet most likely to live?” The Swedish translation of the English “bird with webbed feet” into “bird with swimming feet” provides a much stronger clue and thus a direct translation would have given the Swedish test-takers an unfair advantage. Further, if the original language version does not have an equivalent term for the translated version, psychometric properties or constructs could be lost in the translation (Nasser, 2005).

Brislin, Lonner, and Thorndike recommend that a multiple translation method be used to ensure semantic equivalence (as cited in Beck et al., 2003). The AERA, APA, and NCME (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999) have also subsequently recommended that back-translation should not be a stand alone procedure as it may provide an artificial similarity of meaning across languages but not be the best version of the new language. They recommend that a more iterative process akin to test development and validation be considered in order to ensure that similar constructs are measured across versions. A process involving successive iterations of forward translations and revisions thereof would work equally well, according to Stansfield and Bowles (2007), and the translation equivalence of the two versions can be more cost effective and can take less time.

Bracken and Barona (1991) recommend that researchers use a bilingual committee of judges to compare the original or back-translated version with the translated version of the measure to ensure that the translation is appropriate for the test-takers. Van de Vijver and Hambleton (1996) suggest that the team of translators should have combined expertise in psychology and linguistics. Beck et al. (2003) used multiple translation methods to help ensure the semantic equivalence in translating the Postpartum Depression Screening Scale (PDSS) into Spanish. The multiple methods they employed included back-translation, the committee approach, pretest techniques, and alternate forms equivalence.

The primary concern of translating any instrument is to produce a version that is both linguistically and culturally equivalent to the original. Having reviewed most of the translation techniques, it should be noted that there will always be concepts that cannot be translated into certain languages or that cannot be asked in a meaningful way in certain cultures (Onkvisit & Shaw, 2004).

4.7.2 Translation procedure.

The translation procedure chosen will depend on whether a new instrument is being developed or whether an existing instrument is being translated for a multilingual context. The former is known as simultaneous development while the latter is referred to as successive development. From a methodological perspective, the first option is easier as difficult items such as local idioms which are often difficult to translate can often be avoided (Van de Vijver & Tanzer, 2004). However, in developing countries such as

South Africa the cost of test development necessitates the use of existing instruments. Three options are available to researchers in the successive development method, namely, application, adaptation, and assembly (Van de Vijver & Leung, 1997a, 1997b; Van de Vijver & Tanzer, 2004). These are outlined below.

4.7.2.1 Application.

This option entails the literal translation of an instrument into a target language and it implicitly assumes that the underlying construct is appropriate for each cultural group. The literal translation is commonly used in test translations. (Van de Vijver & Leung, 1997a, 1997b; Van de Vijver & Tanzer, 2004).

4.7.2.2 Adaptation.

This option entails the literal translation of part of the items, changes in other items, or the creation of new items – or any combination of these. For some instruments, it is unrealistic to assume that simple translation would yield construct equivalence for the target cultural group. For example, a measure of anxiety may contain items that need to be reworded to ensure culturally idiosyncratic expressions of the construct are included. Adaptations are based on the premise that using the literal translation would yield a biased instrument. The State-Trait Anxiety Inventory has been adapted into 40 languages, and this approach was used to ensure that the underlying constructs – state and trait

anxiety – were equivalent across language groups (Van de Vijver & Leung, 1997a, 1997b; Van de Vijver & Tanzer, 2004).

4.7.2.3 *Assembly.*

This option entails the adaptation of an instrument to such an extent that it is practically a new instrument. This option is used when construct bias, caused by differential appropriateness of the item content for the majority of the items, threatens a direct comparison. Another reason for using this option would be an incomplete overlap of the construct definition across cultures. For example, aspects of the construct that are salient for some cultures are not covered by the instrument. Researchers contend that Western personality instruments do not cover the personality constructs of other cultures, such as the Filipino and Chinese cultures (e.g. Cheung et al. and Church as cited in Van de Vijver & Tanzer, 2004, p. 123).

4.8 Conclusion

This chapter provided an overview of cross-cultural assessment and the factors that influence cross-cultural assessment. Both bias and equivalence are pivotal concepts that need to be considered in cross-cultural assessment as they help to determine the cultural appropriateness of an instrument. The types of bias and equivalence and how they impact on test scores across different cultural groups as well as how they influence the translation and adaptation of measures were outlined. Guidelines have been provided by

the ITC which have become the benchmark for cross-cultural test adaptation. These were summarized along with the steps that should be addressed when adapting or translating any assessment instrument. A number of techniques have been developed for translating, adapting, and re-norming psychological instruments for cultures and languages other than their initial target population. The techniques that researchers and test developers use in multilingual studies were addressed. The following chapter outlines some cultural approaches to the understanding of childbirth and related mental disorders as well as how these impact on adapting a postpartum depression screening measure cross-culturally.