

# **Principles of Test Development in Papua New Guinea**

**Sarah Flint**

(Sarah.Flint@anu.edu.au)

**Kristen Pammer**

Research School of Psychology, Australian National University

## **Abstract**

With over 800 languages and more than 1000 cultural groups, Papua New Guinea (PNG) provides a challenging environment in which to develop culturally appropriate psychometric tests. Consequently, few tests have been specifically developed for parallel-emic research in PNG. This paper proposes a framework in which to develop psychometric tests within PNG. Linguistic, cultural and social factors are all addressed and strategies for working within these cultural boundaries are posited. Models for translation and validation are assessed in light of the unique challenges presented by the linguistic diversity of PNG. An alternative methodology of translation more appropriate for PNG is also proposed. Furthermore this paper provides a working example of these test development principles. A reading ability test was successfully constructed in line with the principles of test development proposed here. The application of these principles to other Melanesian countries is also discussed.

## **Introduction**

In current psychological research practices, there is a growing acceptance that cultural considerations need to be addressed when undertaking any psychological research. Cognitive psychology, in particular, has been criticized for its reliance on undergraduate university students as participants, thus research is often representative only of western, industrialized and wealthy groups. Awareness of this bias has increased recently and attempts are being made to close this gap (Baddeley, Gardner, & Grantham-McGregor, 1995; Dornyei & Kátona, 1992; Heinrich, Heine, & Norenzayan, 2010). Conducting research with individuals outside of the Western, Educated, Industrialized, Rich and Democratic (WEIRD; Heinrich, Heine & Norenzayan, 2010) nations is vitally important to provide a complete understanding of human cognition.

Of primary concern when undertaking research in a developing country is to determine the validity of the test to be used within that society. This means taking into account linguistic, social and cultural factors. Even in countries that are becoming “westernized” it is necessary to address these factors within each unique cultural setting. Basic health issues such as poor nutrition and limited access to medical care need to be taken into account when assessing cognitive function (Crnic, 1983; Scrimshaw, 1998; Strickland, 2002). For example, adequate protein, especially during childhood is vital for neuronal myelination which is necessary for optimal cognitive functioning. Similarly, limited access to healthcare often means that preventable or treatable illnesses, such as malaria and tuberculosis, as well as parasites such as *Trichuris trichiura* (whipworm) have a negative impact on health, development and cognitive functioning (Connolly & Kvalsvig, 1993; Nokes et al., 1992). Factors which are barely considered in research in developed countries may be important when researching in developing countries. For example individuals born in western countries take for granted that they know their date of birth, however in countries where record keeping is poor or where many births occur outside of a hospital, date of birth or even year of birth may be unknown. These are the types of complications which are apparent when conducting research in Papua New Guinea (PNG).

While this paper addresses PNG in particular, the principles outlined here are applicable to many other countries and cultures. For example, many Melanesian countries share similar cultural practices and nuances as PNG. Likewise, due to their proximity and immigration rates, some remote communities in Northern Queensland, Australia are influenced by similar cultural and linguistic practices as those found in PNG. Con-

sequently, while this paper focuses on PNG and addresses the principles of test development in that country specifically, they are not limited to PNG alone.

There are a number of factors which must be taken into consideration when developing any assessment tool. Bias in test development has been discussed extensively in the last few decades (Hambleton, Merenda, & Spielberger, 2005; Hambleton, Yu, & Slater, 1999; Harkness, van de Vijver, & Mohler, 2003; Matsumoto & van de Vijver, 2012; Poortinga & van de Vijver, 1987; van de Vijver & Leung, 1997, 2000; van de Vijver & Poortinga, 2005; van de Vijver & Tanzer, 1997). Bias refers to differences in test items which do not demonstrate the same meaning within and across cultures (Matsumoto & van de Vijver, 2012). Bias has been discussed on three distinct levels: construct bias, method bias and item bias (Matsumoto & van de Vijver, 2012; van de Vijver & Hambleton, 1996). The importance of understanding bias and controlling for it when adapting or constructing cross-cultural tests cannot be overstated, however as these issues have been discussed extensively elsewhere, the current paper will not focus on these biases. Item format and use of images or words plays an important role in developing an assessment tool, and these have, likewise, been discussed at length in other places (Anderson & Morgan, 2008; Baddeley et al., 1995; Beardsall, 1998; Geisinger, 1994; Goodman, 2005; Irvine & Carroll, 1980; Rosselli & Ardila, 2003). There are many other factors that must be considered when developing a cross-cultural assessment tool, especially in a country as complex as PNG. These factors include choosing the most appropriate approach to undertaking cross-cultural research, etic and emic factors (Berry, 1969), appropriate methodology for translation and development in a pidgin or a creole (Seigel, 1998, 2007), as well as cultural and functional equivalence of words and concepts. From a practical administration standpoint, a researcher must also address medical issues, socioeconomic issues as well as differences in cultural values. These differences in cultural values may include differences in the concept of time and urgency as well as the drive to give socially acceptable answers (see Figure 1 for a flow chart of these considerations). This paper will discuss these factors in principle as they pertain to the development of an assessment tool in PNG.

### **Approaches to Cross-Cultural Research**

There have been numerous papers which have discussed the appropriateness of using different methodologies when conducting cross-cultural research (Harkness et al., 2003; He & van de Vijver, 2012; Van de Vijver & Hambleton, 1996; van de Vijver & Leung, 1997, 2010). Identifying the most appropriate approach to take to conducting cross-cultural research allows for increased validity of the comparison and reduces biases. These approaches have been identified in the literature as adoption, adaptation and assembly. Adoption provides a close translation of an existing instrument for use in a target culture. This method is easy and cheap to implement, as it does not require much else beyond providing the best translation possible. As the adopted test has not been substantially changed it has high validity and provides the ability to compare results across cultures easily. Adaptation is a hybrid method. It allows for a close translation of some items and changes only some of the items. Where linguistic, cultural or psychometric problems render a close translation inappropriate, the item is modified so that it meets the required standard to make it a suitable substitution (He & van de Vijver, 2012). Adaptation is the most commonly used approach when conducting psychometric tests (Hambleton et al., 2005). Assembly refers to the compilation of a new instrument. This methodology maximizes cultural appropriateness as it allows items to be developed which meet all the linguistic, cultural and psychometric requirements. However, this methodology has its limitations. It can be expensive and requires a great deal of time to develop, pilot test, refine and implement. There are benefits and pitfalls to each of the described methodological approaches. It may come down to a trade-off between ecological validity, which is higher in cases of adaptation and assembly, or statistical comparability, for which adoption may provide the best approach (He & van de Vijver, 2012).

Berry, Poortinga, Segall and Dasen (1992) posited that there were two primary and two secondary approaches to cross-cultural research. The primary approaches are etic and emic approaches. Further to this, he posited that secondary approaches further refine the primary approaches. The etic approach to cross-cultural psychology attempts to view a culture from a meta-cultural perspective (Helfrich, 1999). This means that researchers taking this approach use a descriptive system which is deemed equally valid to all cultures, thus does

not take unique cultural factors into account. The objects of comparison, that is the cultures being examined or factors thereof, are operationalized into variables which can be evaluated in a more quantifiable way. Consequently, the comparisons are not aimed at explaining the nature of the culture, rather they aim to examine the influence culture has on thought and behavior. Hence culture is seen primarily as a method of influencing cognition, learning and behavior in individuals who identify with that culture (Berry et al., 1992). The emic approach views culture as an integral part of an individual's identity, not simply a factor which exerts its influence on the individual. Thus the cognition, learning and behavior of an individual cannot be separated from their cultural context and must be examined as a whole cohesive unit.

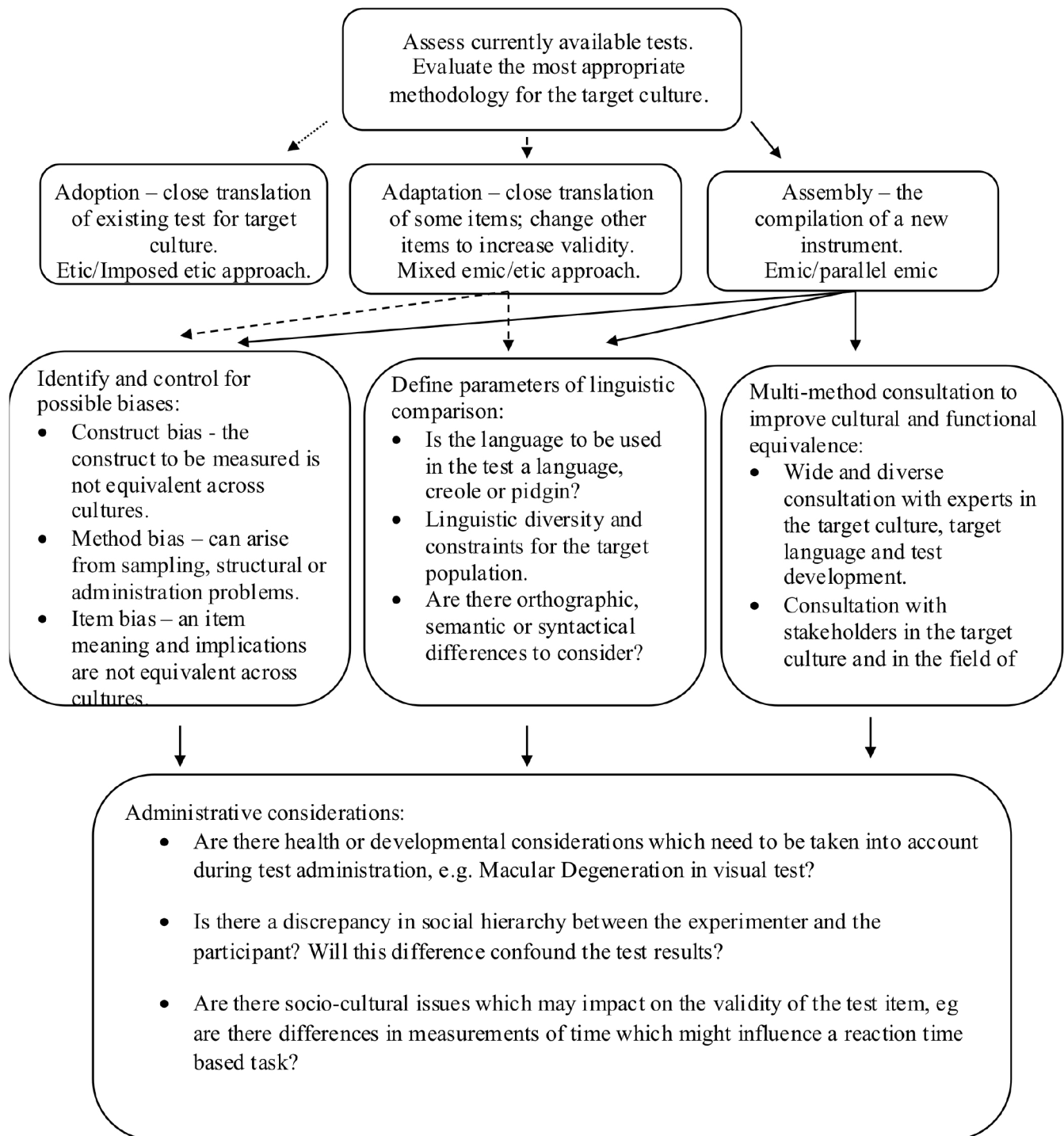
When studying language and culture it is imprecise to view etic and emic concepts as diametrically opposed concepts. Rather, both provide important and complementary information (Berry, 1969, 1989; Niblo & Jackson, 2004). Individually these concepts provide only some of the information necessary to understand the behavior and cognition of individuals, however together the picture becomes a complete view of the drives of an individual within a particular culture. In order to combine these concepts the two secondary approaches can be used. These are imposed etic and parallel emic approaches.

Cross-cultural psychological research has dealt with the problem of conducting research in other cultures in several ways. Imposed etic (Berry, 1969) research is where researchers take a Western developed and standardized test and apply it directly to another culture. This idea relates to the use of an adopted test methodology for psychometric tests (He & van de Vijver, 2012; van de Vijver & Leung, 2010). Understandably this has led to many researchers finding a difference between cultures on the particular test, however the validity of this approach has been questioned (Smith & Bond, 1993) as this methodology is based on the faulty assumption that the items in the test would have the same meaning to the test participants as they had in their original culture. There are many examples of where this is not the case. Take for example the Western concept of family which generally means nuclear family, but may be stretched to include blood relatives. In Papua New Guinea, the concept of family can extend to an entire village, and in a modern context may also include close friends who have been given the same rights as family members.

An alternative methodology, termed parallel emic studies (Berry, 1969, 1989), requires the construction and validation of an assessment tool within the culture it is designed to be tested, as in the case of the assembly methodology (He & van de Vijver, 2012; van de Vijver & Leung, 2010) discussed earlier. This is a much more complicated process. It requires the test developer to have a near-native understanding of cultural, linguistic and social components of the culture in order to accurately represent these in the assessment. Parallel emic assessments could be used for cross-cultural assessments only once the tool has been deemed valid within the culture in which it was written. When the tool has been tested across cultures, cautious comments can be made about the universality of some of these features. Testing across cultures should be done only with extreme caution. Creating two parallel emic assessments in two different cultures and then attempting to validate the comparison is problematic although not impossible. When an assessment tool has reached this stage it can be regarded as derived-etic.

From this discussion it is evident that imposed etic is the least desirable research methodology (Berry, 1969, 1989, 1999). An assessment developed with an imposed etic approach cannot be deemed adequate as there are inherent cultural peculiarities which need to be considered in each cultural context. While there are both advantages and disadvantages to the application of etic, parallel emic and derived-etic and adaptation and assembly methodologies, the complexity associated with cultural and linguistic factors in PNG means that an assembly based parallel emic approach will provide the best cultural validity. Parallel emic, and by extension derived-etic, should be the 'gold standard' for cross-cultural research to ensure as many cultural confounds are taken into account as possible. Despite the long standing evidence for the lack of validity in the imposed etic model, a meta-analysis of cross-cultural research found that of 635 empirical articles, 591 (or 93%) were imposed etic (Ongel & Smith, 1994). Furthermore, only 6% could be considered emic and less than 1% met the criteria for derived-etic. In a more recent review of articles published by the Australian Psychological Society in the 20 years preceding 2004 (Niblo & Jackson, 2004), it was found that of the 35 articles reviewed, 89%

(or 31 articles) were imposed etic, one article was emic, and three articles (9%) were derived-etic. This over-reliance on imposed-etic methodology questions the validity of the findings of many cross-cultural research articles. Parallel emic research is the most desirable methodology. This is all the more apparent when research methodology is applied to cultures outside the experience or understanding of the researcher, for example when Western researchers undertake research in PNG.



**Figure 1.** Flow chart of the principles involved in test development in Papua New Guinea and considerations which are recommended.

Papua New Guinea is a culturally rich country with many facets to make it ideal for psychological research. When researching in PNG the parallel emic version of test development is certainly preferred. Papua New Guineans have a long history of ‘fear of the other’ which can impede research undertaken there. This



impediment is evinced through, for example, lack of support from leaders or elders, which can disrupt or draw out the approval process, or participants supplying what they believe to be the desired answer, as in the case of Margaret Meads research in PNG. This is fear of the other not without justification. For decades Papua New Guinea was colonized or administered by different countries. Prior to Australia's administration of PNG after the First World War, the British controlled the southern area called British New Guinea and the Germans controlled the northern half of the island, known as Papua. During the period of colonization by the British and Germans, Papua New Guineans were more or less forced into an indenture system of labor. This remained more or less unchanged under the Australian administration of PNG (Fitzpatrick, 1980). During this time Papua New Guinean workers were considered to be little better than slaves (Fitzpatrick, 1980; Nibbrig, 1992). Given this history it is not surprising that there is a reluctance to engage with people from outside PNG. Therefore, in order to successfully conduct research in PNG a parallel emic approach, which is sensitive to the culture and history of PNG, can reduce the fear of the other ingrained in some Papua New Guineans.

However this xenophobia runs deeper than simply distrust of non-Papua New Guineans. Papua New Guinea has itself only been an independent country since 1992 when it was granted independence from Australian administration. Since independence, Papua New Guinea has sought to become a unified country from what was, even during colonization, a diverse and at times hostile collection of cultural and ethnic groups. For example, in the Southern Highlands Province tribal violence is a significant problem (Kopi, Hinton, & Robinson, 2010). In 1996 24% of all trauma admissions to Mendi hospital were related to tribal violence (Matthew, Kapua, Soaki, & Watters, 1996). Tribal violence is often carried out in rural areas between culturally and geographically similar people (Macintyre, 2008).

This violence is generally the result of a number of social and economic factors as well as historical precedents. While this type of violence is not usually directed to Westerners, when conducting research in PNG it is important to understand the fragile nature of social equity in PNG. For example, when constructing assessment materials it is important to acknowledge that even to the level of clothing there are cultural differences in different areas of PNG. Men in the Southern Highlands find it acceptable to grow a beard at any time, whereas men in the coastal regions grow a beard only when they are in mourning. An understanding of the sociocultural issues that may arise in PNG is needed so these differences can be taken into account in illustrations and discussions. This will help avoid giving offence, distracting from the property being tested, and provide an accurate assessment that is not unduly biased.

### **Language Use in Test Development**

Papua New Guinea is a developing country with a population of almost 5.2 million people (National Statistical Office of Papua New Guinea, 2010) who speak over 800 languages (Lewis, 2009). PNG has three official languages; Motu (the lingua franca of the people from the Papuan region), Tok Pisin (Pidgin) and English. The most widely spoken of these languages are Tok Pisin and English. The majority of people in PNG, particularly in areas that are close to main cities and towns, can speak their own language, Tok Ples, as well Tok Pisin and/or English. Individuals who are themselves English speakers tend work in government and related areas, the service industry or hospitality. Beyond early primary school, English is the primary language used in classrooms.

Accurate and appropriate language use is vital to the development of valid assessment tools. There are many factors to address in relation to linguistic usage including the nature and development of the language to be used in an assembled test. For example, establishing the developmental level of the language in question will determine how it is structured and later can be translated. Understanding whether the language is a pidgin, creole or fully developed language is important. A pidgin or a creole cannot be treated the same way as a fully developed language can, as they do not have the structure or regularity that a fully developed language does. As PNG is a linguistically diverse country and many people within PNG are bilingual or are polyglots, it is important to address how this linguistic diversity can influence the development of psychometric tests. Similarly, functional and cultural equivalence provides an important footing for linguistic comparison. Likewise, orthographic regularity can help determine how best to evaluate a linguistic segment and compare across segments.

*Pidgin and creole assessment and translation.* In order to accurately develop an assessment tool in a creole language it is first important to understand the origins of the creole and the origins of the pidgin from which it has developed. A pidgin is considered a contact language. This means that it arose, often spontaneously, as a mixture of other languages as a means of communication between speakers of different languages. A pidgin is characterized by a limited vocabulary and a simple grammatical structure and is learnt only as a second language. In PNG Tok Pisin started out as a pidgin language. It developed as a means by which a master could talk to his workers, and amongst the workers themselves who had often come from different cultural and linguistic backgrounds.

A creole has a basic and stable structure. It is starting to have a unique vocabulary and can be learnt as a primary language. All creoles develop from pidgin languages. In PNG, Tok Pisin is now considered a creole. While there are variations across PNG in pronunciation and slang, there is a more or less stable grammatical structure and accepted basic vocabulary. Due to the transient and continually evolving nature of a pidgin, and to a lesser extent a creole, it can be very difficult to use conventional translation methods to translate to and from these languages.

Brislin (1970, 1986) proposed one of the most commonly used translation methods, the translation -- back translation model. In this translation method each item is translated from the source language (SL) into the target language (TL) by a speaker fluent in both the SL and the TL. Another speaker, also fluent in the SL and the TL then takes the translation, now in the TL and translates it back into the SL. The original and adapted versions of the test can be reviewed with a view to resolving issues in semantic, idiomatic, experimental and conceptual equivalence. A pre-test is then recommended using judgmental procedures. This methodology can be expanded further to include field-tests and full pilot studies to increase the validity of the test and identify any discrepancies which may have gone unnoticed in judgmental procedures (Hambleton & Patsula, 1998; van de Vijver & Hambleton, 1996).

While this is a functional method for translation in fully developed and stable languages, although it should be noted that linguistic equivalence is not always possible due to the isomorphic nature of language (Harkness & Schoua-Glusberg, 1998), this methodology is not suited to rapidly evolving languages such as pidgins and creoles. In Tok Pisin a more appropriate methodology is to gather a number of translations from the source language to the target language. Given the diversity in the way the language is operationalized, gathering translations from several individuals from different cultural and linguistic backgrounds would increase the accuracy and applicability of the translations. This 'team approach' is advocated by recent research in adapting and translating test questions (Harkness, 2008; Harkness & Schoua-Glusberg, 1998). A speaker fluent in both the target and source languages (Harkness, 2008; Harkness et al., 2003) is then able to use judgmental procedures to assess the translations and the 'most correct' translation identified and used (Hambleton & Patsula, 1998). The purpose of this population-based methodology is to identify possible differences in translations and assess which can be most readily used and provide the most reliable result. The translation -- back translation method does not make allowances for the scope of differences found in pidgin and creole translation, as in this methodology, differences compound across the forward and back translations rather than resolving. The population-based methodology identifies these differences and resolves them using expertise in both languages. It is still recommended that field testing and pilot studies are undertaken, as suggested by the International Test Commission (2001), to ensure validity and reliability (Hambleton, 2001; Hambleton et al., 2005; Hambleton et al., 1999; Tanzer & Sim, 1999; van de Vijver & Hambleton, 1996; van de Vijver & Leung, 1997; van de Vijver & Poortinga, 2005).

*Linguistic diversity.* When developing a test in a country as linguistically diverse as PNG the researcher must first define the scope of the linguistic aspect of the test. With over 800 languages, many of them only oral languages, it may be unrealistic to assemble a linguistically equivalent test for each language. Even taking into account that there are three official languages in PNG, the use of all of these languages may not be appropriate. Motu, for example, is spoken only by people from the Papuan region of PNG, thus if the test is to be administered outside of this region a Motuan translation would be redundant. Language selection in PNG will depend on the region in which the testing is being conducted, as well as the demographics of the participants. For

example, conducting research in one of the settlements in Port Moresby is not the same as conducting research in one of the traditional villages in and around Port Moresby. Despite being located in the Papuan region, most of the people living in the settlements have come from other regions in PNG and so few speak Motu, in contrast to the traditional villages where Motu is the most commonly spoken language. Judgment on the part of the researcher in consultation with individuals who understand the linguistic constraints of the people for whom the test is being developed will provide the most reliable linguistic foundation for test development.

*Cultural and functional equivalence.* It is important to control for cultural and functional equivalence when conducting cross-cultural research. A concept in one culture is not necessarily directly transferable to another culture. For example in western cultures the concept of family is generally understood to refer to a nuclear family, and possibly including grandparents, but not necessarily uncles and aunts and cousins. While in Papua New Guinean Tok Pisin there are a number of different words for different concepts of 'family'. *Famili* refers to a nuclear family in the same way as family is commonly used in English. *Lain* refers to an extended family, which may include up to second and third relations, or in fact a whole village. Finally, *wantok* is used in the context of a cultural group, for example Motuan people (the traditional landowners of the area around Port Moresby), or a social group, for example close friends and family. Specificity when conducting psychological research is important, not only to gain the most accurate information, but also to be able to standardize responses. Therefore, understanding cultural and functional equivalence can prevent misunderstandings. Following on from the example above, there would be a difference in the response given by English speakers who were asked questions about their family, where 'family' is understood to refer to nuclear family, compared to Tok Pisin speakers who were asked about their '*wantok*', an acceptable translation for family but with different connotations. With such a diverse range of meanings and specific words for each context it is important to be appropriate when employing a particular word.

To ensure that there is consistency in concepts and syntax, the development of these tests would be undertaken by individuals who have a deep understanding of the complex nature of culturally diverse test development. For researchers, especially those who are not native to Papua New Guinea, a wide consultation process, both formally and informally in manner, may be utilized to establish the use of appropriate words and concepts and increase the validity of the measures. Such a procedure has been discussed for use in Sub-Saharan Africa (Abubakar, van de Vijver, van Baar, Kitsao-Wekulo, & Holding, 2009). Abubakar and colleagues (2009) advocated for a multimethod consultation approach, including a diverse range of consultation methods and attention being given to the appropriateness of the consultants chosen as cultural understanding of the target population can help ensure the validity of the research.

*Orthographically regular and irregular language differences.* Orthographic regularity refers to the consistency with which graphemes correspond to phonemes. In the case of orthographically regular languages, words are spelt as they sound. Orthographical regularity plays a significant role in the processing speed and cognitive load that reading text places on an individual. Research into individuals with dyslexia and other Specific Language Impairments (SLI) has demonstrated that reading speed in orthographically regular languages, such as German and Italian, is faster when compared to orthographically irregular languages such as English. In one study (Landerl, Wimmer, & Frith, 1997), German and English dyslexic readers were compared on single word/non-word reading and phonological processing tasks. They found that German dyslexic readers were faster at both tasks than their English counterparts. Landerl, Wimmer and Frith posited that this difference was due to the orthographic differences of the languages (1997). Orthographically regular languages are easier to process than orthographically irregular languages. Numerous other studies have supported the idea that orthography plays an important role in reading speed (Di Filippo et al., 2005, 2006; Frith, Wimmer, & Landerl, 1998; Hutzler & Wimmer, 2004; Paulesu et al., 2000).

This research has important implications for bilingual research in PNG. Creoles are generally orthographically regular languages as they have evolved from pidgins which start as spoken language, where the written form comes after the pidgin is established. Consequently, Tok Pisin is an orthographically regular language. This difference is important particularly when conducting cross cultural research where one language,

for example English, is orthographically irregular language and the other, such as Tok Pisin, is orthographically regular. In cases such as this where there is a discrepancy in orthographic regularity it is necessary to be selective about outcome measures which can be applied. For example, a timed reading test or a test of reading speed would not provide a meaningful measure of reading ability. It is easier and faster to read orthographically regular words than to read orthographically irregular word (Landerl, Wimmer & Frith, 1997). Thus a reading speed test in this context would provide only a measure of the orthographical differences between the two languages.

### **Administrative Considerations**

Alongside linguistic and cultural considerations, medical, socioeconomic and differences in cultural values, such as understanding time and urgency, may be important for accurate test development. Papua New Guinea faces many challenges with developing nationhood. Access to education for the whole population is certainly a major consideration, but it is mediated by infrastructure, financial and health constraints. Test items and administration techniques should be sensitive to these factors. In some cases it may be necessary to administer additional tests prior to conducting the primary research. For example, due to the cost and limits in accessing visual health care it may be important to administer visual tests, such as the Amsler grid test for Macular Degeneration (MD) or the Landolts C test for visual acuity, prior to reading test. While official statistics are not available for rates of MD in the general PNG population, anecdotal evidence shows that MD rates are much higher than expected in young adults, although this varies between villages. In PNG there are cases of MD in individuals in their 20s, where as in Australia it is most common for MD to be related to aging. In Australia in 2004 the most common cause of blindness was age-related MD (48%; Taylor et al. 2005).

Papua New Guinea is a country which values respect and has a clear social hierarchy. Landowners are generally the wealthiest people in a village and also wield the greatest social and political influence within that village. Landowners are treated with respect and deference, such as is evident towards politicians and business owners. Education is highly valued by Papua New Guineans, particularly by families who can barely afford to send their children to school (Papua New Guinea Education Advocacy Network, 2011). Consequently, educated individuals are also held in high esteem. When asked a question or request by someone who is deemed to be of a high social standing, it is common practice in PNG to do as much as possible to assist this person. This can be taken to the point of agreeing to something that is impossible to do so that the person of high standing does not lose face.

When conducting research in PNG, especially as educated, relatively wealthy person, it is important to phrase questions and design tests to put the participant at ease and in a position of as close to equality as possible. This is especially important when the researcher is a non- Papua New Guinean.

Our understanding of time is a social and cultural construct (Arrow, Poole, Henry, Wheelan, & Moreland, 2004). Time can be discussed in several ways: clock time; predictable time (such as the predictable cycle of the seasons); goal motivated time; unpredictable event and predictable event time (Ancona, Okhuysen, & Perlow, 2001). For the purpose of this discussion we will be focusing on the notions of event time. While it is generally accepted that time can be measured within the 24 hour day, the standard of time used in relation to events is less certain. In most western countries, when a meeting time, for example, is agreed upon all the meeting participants will turn up within minutes of the appointed time. In PNG, however, this is not the case. Adherence to a strict idea of event times is uncommon in PNG (Mellam, 2002; Mellam & Espnes, 2003). Time is a fluid concept, so a person could turn up an hour late for a meeting and this would not be socially unacceptable. Within the context of undertaking research in PNG, it is important to remember that simply because you arranged a time to meet with a participant or contact there is no certainty that that meeting will take place at that time. Likewise, when conducting research which is based on reaction times, it is important to note that the reaction times of a PNG participant may be quite different to the reaction times of an Australian participant, for example. This difference in reaction time makes reaction time based tests difficult to administer and measure. For an Australian participant a task which measures reaction time seeks to gain a 'first response' to the stimuli by reducing time to think about the response. In PNG this concept is outside of the experience of many Papua New Guineans, consequently due to habit or to difficulty in responding in the desired way their



reaction times may be slower than Australian participants. Because the difference in reaction times may be culturally influenced it may not be accurate to compare reaction times without moderating the results through an understanding of these cultural differences.

### Conclusion

Test development in PNG not a straightforward process. In order to develop the most accurate tool possible, it is necessary to evaluate the most appropriate methodology and evaluate this in light of etic and emic factors. The gold standard for cross-cultural research in PNG should be the assembly approach which utilizes parallel-emic methodology. Using parallel-emic methodology will reduce the number of cultural value and cultural prejudice related confounds. Orthographic and semantic difference must be considered seriously and weighed against the possible benefits of using components of a test with such confounds. Translation method is another very important consideration. There may be times when conventional back-translation methods are most appropriate, however should be used with caution for creole languages. In such instances a moderated translation approach may be more appropriate. Finally, there are numerous socioeconomic factors specific to the culture which must at least be acknowledged and ideally controlled for when developing a cross-cultural test in Papua New Guinea.

### Author Note

The authors would like to acknowledge the assistance of the Papua New Guinea National Research Institute, and the Papua New Guinea Education Advocacy Network.

### References

- Abubakar, A., van de Vijver, F. J. R., van Baar, A., Kitsao-Wekulo, H., & Holding, P. (2009). Enhancing psychological assessment in Sub-Saharan Africa through participant consultation. In A. Gari & K. Mylonas (Eds.), *Quod erat demonstrandum: from Herodotus' ethnographic journeys to cross-cultural research*, (pp. 169-178). Athens, Greece: Pedio Book Publishing.
- Ancona, D. G., Okhuysen, G. A., & Perlow, L. A. (2001). Taking time to integrate temporal research. *The Academy of Management Review*, 26, 512-529.
- Anderson, P., & Morgan, G. (2008). *Developing tests and questionnaires for a national assessment of educational achievement*. Washington, DC: The World Bank.
- Arrow, H., Poole, M. S., Henry, K. B., Wheelan, S., & Moreland, R. (2004). Time, change and development, the temporal perspective on groups. *Small Group Research*, 35, 73-105.
- Baddeley, A. D., Gardner, J. M., & Grantham-McGregor, S. (1995). Cross-cultural cognition: Developing tests for developing countries. *Applied Cognitive Psychology*, 9, S173- S195.
- Beardsall, L. (1998). Development of the Cambridge Contextual Reading Test for improving the estimation of premorbid verbal intelligence in older persons with dementia. *The British Journal of Clinical Psychology*, 37, 229-240.
- Berry, J. W. (1969). On cross-cultural comparability. *International Journal of Psychology*, 4, 119-128.
- Berry, J. W. (1989). Imposed etics - emics - derived etics: The operationalisation of compelling ideas. *International Journal of Psychology*, 24, 721-735.
- Berry, J. W. (1999). Emics and Etics: A symbiotic conception. *Culture Psychology*, 5, 165- 172.
- Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (1992). *Cross-cultural psychology: Research and applications*. Cambridge University Press.
- Brislin, R. W. (1970). Back translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1, 185-216.
- Brislin, R. W. (1986). Wording and translation of research instruments. In W. Lonner & J. W. Berry (Eds.). *Field methods in cross-cultural research*. (pp. 137-164). Thousand Oaks, CA: Sage Publications, Inc.
- Connolly, K. J., & Kvalsvig, J. D. (1993). Infection, nutrition and cognitive performance in children. *Parasitology*, 107, S187-S200.
- Crnac, L. S. (1983). Effects of Nutrition and environment on brain biochemistry and behavior. *Developmental Psychobiology*, 16, 129-145

- Di Filippo, G., Brizzolara, D., Chilosi, A., De Luca, M., Judica, A., Pecini, C., et al. (2005). Rapid naming, not cancellation speed or articulation rate, predicts reading in an orthographically regular language (Italian). *Child Neuropsychology*, *11*, 349-361.
- Di Filippo, G., Brizzolara, D., Chilosi, A., De Luca, M., Judica, A., Pecini, C., et al. (2006). Naming speed and visual search deficits in readers with disabilities: Evidence from an orthographically regular language (Italian). *Developmental Neuropsychology*, *30*, 885-904.
- Dornyei, Z., & Katona, L. (1992). Validation of the C-test amongst Hungarian EFL learners. *Language Testing*, *9*, 187-206.
- Fitzpatrick, P. (1980). Really rather like slavery: Law and labor in the colonial economy of Papua New Guinea. *Crime, Law and Social Change*, *4*, 77-95.
- Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading*, *2*, 31-54.
- Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, *6*, 304-312.
- Goodman, K. (2005). The Perfect Literacy Test. *Language*, *5*, 24-27.
- Hambleton, R. K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment*, *17*, 164-172.
- Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Erlbaum.
- Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research*, *45*, 153-171.
- Hambleton, R. K., Yu, J., & Slater, S. C. (1999). Field-test of the ITC Guidelines for adapting educational and psychological tests. *European Journal of Psychological Assessment*, *15*, 270-276.
- Harkness, J. A. (2008). Comparative survey research: Goals and challenges. In E. de Leeuw, J. Hox & D. Dillman (Eds.), *International handbook of survey methodology* (Vol. 1, pp. 56-77). New York, NY: Erlbaum.
- Harkness, J. A., & Schoua-Glusberg, A. (1998). Questionnaires in translation. *ZUMA Nachrichten Spezial*, *3*, 87-127.
- Harkness, J. A., van de Vijver, F. J. R., & Mohler, P. P., (2003). *Cross-cultural survey methods*. Chichester, NJ: Wiley.
- He, J., & van de Vijver, F. J. R. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture*, *2*(2).
- Heinrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavior and Brain Sciences*, *33*, 61-135.
- Helfrich, H. (1999). Beyond the dilemma of cross-cultural psychology: Resolving the tension between etic and emic approaches. *Culture & Psychology*, *5*, 131-153.
- Hutzler, F., & Wimmer, H. (2004). Eye movements of dyslexic children when reading in a regular orthography. *Brain and Language*, *89*, 235-242.
- International Test Commission. (2001). International guidelines for test use. *International Journal of Testing*, *1*, 93-114.
- Irvine, S. H., & Carroll, W. K. (1980). Testing and assessment across cultures: Issues in methodology and theory. In H. Triandis & J. Berry (Eds.), *Handbook of cross-cultural psychology* (Vol. 2: Methodology, pp. 181-244). Boston, MA: Allyn & Bacon.
- Kopi, M., Hinton, R., & Robinson, S. (2010). Violence and Insecurity in the Southern Highlands of Papua New Guinea. Retrieved from <http://www.oxfam.org.nz/report/violence-and-insecurity-in-the-southern-highlands-of-papua-new-guinea>
- Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition*, *63*, 315-334.
- Lewis, M. P. (2009). Ethnologue: Languages of the world sixteenth edition. Retrieved from <http://www.ethnologue.com/>
- Macintyre, M. (2008). Police and thieves, gunmen and drunks: Problems with men and problems with society in Papua New Guinea. *The Australian Journal of Anthropology*, *19*, 179-193.
- Matsumoto, D., & van De Vijver, F. J. R. (2012). Cross-cultural research methods. In H. Cooper (Ed.), *APA handbook of research methods in psychology* (Vol. 1.: Foundations, Planning, Measures, and Psychometrics, pp. 85-100).

- Washington, DC: American Psychological Association.
- Matthew, P. K., Kapua, F., Soaki, P. J., & Watters, D. A. K. (1996). Trauma admissions in the southern highlands of Papua New Guinea. *Australian and New Zealand Journal of Surgery*, *66*, 659-663.
- Mellam, A. (2002). Socio-demographic markers on Type A behavior. *International Congress Series*, *1241*, 105-118.
- Mellam, A., & Espnes, G. A. (2003). Emotional distress and the Type A behaviour pattern in a sample of civil servants. *Personality and Individual Differences*, *34*, 1319-1325.
- Nibbrig, N. E. H. (1992). Rascals in paradise: Urban gangs in Papua New Guinea. *Pacific Studies*, *15*, 115-134.
- Niblo, D. M., & Jackson, M. S. (2004). Model for combining the qualitative emic approach with the quantitative derived etic approach. *Australian Psychologist*, *39*, 127-133.
- Nokes, C., Grantham-Mcgregor, S., Sawyer, A. W., Cooper, E. S., Robinson, B. A., & Bundy, D. A. P. (1992). Moderate to heavy infections of *Trichuris trichiura* affect cognitive function in Jamaican school children. *Parasitology*, *104*, 539-547.
- Ongel, U., & Smith, P. B. (1994). Who are we and where are we going?: JCCP approaches its 100th issue. *Journal of Cross-Cultural Psychology*, *25*, 225-253.
- Papua New Guinea Education Advocacy Network. (2011). *Survey of education experience*. Port Moresby: Asia-South Pacific Education Watch.
- Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., et al. (2000). A cultural effect on brain function. *Nature Neuroscience*, *3*, 91-96.
- Poortinga, Y. H., & van de Vijver, F. J. R. (1987). Explaining cross-cultural differences bias analysis and beyond. *Journal of Cross-Cultural Psychology*, *18*, 259-282.
- Rosselli, M., & Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: a critical review. *Brain and Cognition*, *52*, 326- 333.
- Scrimshaw, N. S. (1998). Malnutrition, brain development, learning, and behavior. *Nutrition Research*, *18*, 351-379.
- Seigel, J. (1998). Literacy in Melanesian and Australian pidgins and creoles. *English World Wide*, *19*, 105-121.
- Seigel, J. (2007). Literacy in pidgin and creole languages. In A. Liddicoat (Ed.), *Language planning and policy: Issues in language planning and literacy* (pp. 143-163). Clevedon, United Kingdom: Multilingual Matters.
- Smith, P., & Bond, M. (1993). *Social psychology across cultures: Analysis and perspectives*. New York, NY: Harvester Wheatsheaf.
- Strickland, S. S. (2002). Functional consequences of adult malnutrition in developing countries: A review. *Journal of Physiological Anthropology and Applied Human Science*, *21*, 1-9.
- Tanzer, N. K., & Sim, C. Q. (1999). Adapting instruments for use in multiple languages and cultures: A review of the ITC guidelines for test adaptations. *European Journal of Psychological Assessment*, *15*, 258-269.
- Taylor, H. R., Keeffe, J. E., Vu, H. T., Wang, J. J., Rochtchina, E., Pezzullo, M. L., & Mitchell, P. (2005). Vision loss in Australia. *Medical Journal of Australia*, *182*, 565-568.
- van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, *1*, 89-99.
- van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.
- van de Vijver, F. J. R., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology*, *31*, 33-51.
- van de Vijver, F. J. R., & Leung, K. (2010). Equivalence and bias: A review of concepts, models and data analytic procedures. In D. Matsumoto & F. J. R. van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 17-46). Cambridge, United Kingdom: Cambridge University Press.
- van de Vijver, F. J. R., & Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. In R. Hambleton, P. F. Merenda & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 39-63). Mahwah, NJ: Erlbaum.
- van de Vijver, F. J. R., & Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, *47*, 263-280.