



The use of Western Standardized Psychological Tests in Non-Western Contexts

Dylan Marshall

International School Manila, Philippines

Abstract The aim of this literature review is to examine the issues related to the use of standardized psychological tests developed in the West within non-Western contexts. The central thesis is that it is vital to be aware of the cultural differences and how these may affect performance on standardized psychological tests. Psychologists must be aware of this when they are assessing and evaluating students in schools in non-Western cultures. The review first presents a theoretical framework for understanding cultural differences in psychology. It then reviews researches that highlight the methodological and ethical controversies related to the use of standardized tests within non-Western contexts. The review then examines the steps that can be put in place to ensure equivalence when using standardized tests outside of the context that they were developed. This paper concludes by arguing that culturally relevant tests should be developed in a local context.

Keywords: culture, cultural bias, standardized testing, non-western contexts

Psychology has become a globalized discipline but it is important to understand that modern psychology has its roots in Europe and North America, and the majority of research still takes place on these two regions. For example, Smith and Bond (1998) carried out a review of psychology textbooks and found that only 10 per cent of the world's population was represented in samples used by psychological researchers. There is therefore still a Western dominance in psychological research and although this is gradually diminishing it is still a pertinent issue today. The problem arises when psychologists develop theories and models, which they claim to be 'universal' to all human behavior when they are only really applicable to the cultural context in which they were developed.

The cultural context of theories is highly relevant to the development and use of standardized psychological tests in schools. It is of utmost importance for school psychologists to be aware of this because when standardized psychological tests developed in Western countries are used

for evaluation purposes in non-Western countries, unfair and incorrect decisions could be made. In this review standardized tests are defined as tests that was designed in such a way that the questions, conditions for administering, scoring procedures, and interpretations are consistent and applied in a predetermined, standard manner, likewise, it also involves making comparison of the test taker to a set of norms from a particular population (Popham, 1999).

Theoretical framework for the relation between culture and standardized testing

Culture is a very difficult concept to define. Matsumoto (2003), for example, mentions a book from 1998 that analyzed as many as 128 different definitions of culture. One of the most overarching definitions of culture was proposed by Herskovits (1948) who saw culture as the ‘human-made part of the environment’. Culture is a concept that is used to describe many different things, from food and eating habits, clothing and communication patterns to rituals and religion. Triandis (1990) claims that culture has an objective (physical materials) and subjective (cultural norms and beliefs) elements. According to Shaules (2007) the actual culture cannot be seen but it is only its manifestations that can be witnessed. In his view, there is ‘deep culture’ which is related to beliefs, attitudes, and values that underpin cultural manifestations. In addition, Hofstede (2001) describes culture as ‘mental software’, which are cultural schemas that have been internalized and has an influence on thinking, emotions and behavior. According to Hofstede (2001), members of a sociocultural group share this mental software and it is learned through daily interactions and feedback from other members of the group.

Given the definitions of culture, it is reasonable to come to a conclusion that culture affects the way we think and behave to a large extent. However, this factor has often been ignored by psychologists when they carry out their research, and continues to be ignored by psychologists when they administer standardized tests developed in the Western countries to participants from non-Western countries. Considering the issues related to this, test users should consider the conflict between cultural ‘absolutism’ and ‘relativism’. Cultural absolutism is associated with the ‘mainstream’ scientific psychology that has been conducted in North American and European universities during the twentieth century as such, it is no surprise that the major names of standardized tests such as Raven (1936); Otis and Lennon (1957), Flanagan (1960) and Cattell (1973) were all from this tradition while relativism is the approach which has been more central to the field of anthropology. This tradition tends to give more weight to social and cultural factors than to biological ones a direct reverse of the absolutists. Relativists have little interest in studying intergroup similarities while absolutists believe that there is a similarity between groups and in their research they look for ‘the psychic unity of mankind’ (Segall, Dasen, Berry, & Poortinga, 1999). When it comes to psychological testing relativists argue that it is impossible to have context -free categories of measurement. They avoid comparisons whereas absolutists attempt to use context-free measurements and standardized psychological instruments.

Since most of the psychologists who developed the standardized psychological tests were absolutists in their approach they have taken an ‘etic’ approach to understanding human behavior as opposed to an ‘emic’ one. Taking his cue from cultural anthropology, Pike (1954) made an important distinction between emic and etic. Etic are universal behaviors and constructs which all human societies have, such as marriage, concepts of intelligence, time, and the education of children. On the other hand, emic are culture-specific behaviors and constructs. This distinction is of importance when considering the development and use of standardized psychological tests. The Western psychologists who developed these tests thought they were measuring etic that are universal characteristics that are common to all human beings. However, when such tests were carried out in non-Western cultures, it became clear that different people approached problems in different ways, suggesting that they had developed a measure of their own culturally specific emic. The problem arises when these standardized measures developed in the West are directly applied to non-Western cultures. This known as ‘imposed etic’.

Brislin (1993) clearly illustrates the problems that can arise from such an approach. For the concept of intelligence, he suggests that an etic definition of intelligence as the capacity to excel in ‘solving problems, the exact form of which hasn’t been done before’ is a definition likely to be generally acceptable to different cultures. However, intelligence also has emic aspects, the emic of ‘mental quickness’ (for example measured by timed IQ tests) has been found not to be valid in all cultures. For example, for the Baganda people of Uganda, intelligence is associated with slow, careful and deliberate thought (Wober, 1974).

An example of ‘imposed etic’ in psychological research comes from Glick (1975) who wanted to measure intelligence in the Kpelle tribe from rural Liberia, West Africa. Glick asked participants to sort twenty objects into groups. They did this using ‘functional’ grouping (knife with orange, potato with hoe). The experimenter was expecting normal ‘category’ sorting (knife, hoe, potato, orange) and when challenged the participants said ‘that was how the wise man would do it’. When asked how a fool would do it they sorted the objects into the categories the experimenter was originally looking for. These examples show how the use of a standardized test of intelligence such as the Raven’s (1936) Progressive Matrices would not be a relevant or fair measure for the Baganda or Kpelle people, as the test is based on a Western emic concept of intelligence, which is radically different to the Baganda, or Kpelle’s concept of intelligence.

It is also important to consider cultural differences when testing for personality. The most widely accepted model of personality is the Five Factor Model (FFM) developed by Costa and McCrae (1992) which consists of five broad domains of personality which are *openness*, *conscientiousness*, *extraversion*, *agreeableness*, and *neuroticism*, assessed using the Revised NEO Personality Inventory (NEO-PI-R). McCrae and Terracciano (2005) reviewed research carried out in fifty cultures on six different continents and concluded that in most cultures around the world personality traits can be summarized by the FFM. This suggests that FFM is a universal etic construct, present in all cultures. McCrae (2004) states that “trait structure, age and gender differences, and cross-observer agreement are all universal”

(p.3) and Digman and Inouye (1986) argued that the FFM was applicable non-Western cultures such as Japan, China and the Philippines.

However, Guthrie and Bennett (1971) carried out research with Filipino participants and found profound differences in personality structure when comparing Filipino participants with American norms. The Filipino sample had similar conception of *extraversion* to the American sample, but some traits factored down differently to the American sample. For example, the factor of *emotional stability* was split into two factors in the Filipino sample, which were *worry/anxiety* and *somatic symptoms*. Furthermore, McCrae and Terracciano (2005) found that Filipinos, score low on *neuroticism* when compared to other cultures.

There is a debate over whether or not the FFM accurately captures the construct of personality in all cultures. Some alternative indigenous measures of personality have been developed in non-Western countries. For example the Chinese Personality Assessment Inventory (CAPI) includes a trait known as *interpersonal relatedness*, which is not included in the FFM (Funder, 2010). Katigbak, Church, Guanzon-Lapeña, Carlota, & del Pilar, (2002) carried out a study of personality in the Philippines using indigenous Filipino personality scales as well as the NEO-PI-R (Costa & McCrae, 1992). They found that there was an overlap between the Filipino scales and the FFM, but they also found indigenous factors such as *pagkamadaldal* (social curiosity) and *pagkamapagsapalaran* (risk-taking), which do not feature in the FFM.

Even though there is a large amount of research, which supports the FFM's cross-cultural validity, research also highlights the existence of other personality traits apart from those included in the FFM. More research is needed in this field to further our understanding of personality across different cultures. Psychologists should be aware that the intelligence and personality tests they use might not be able to fully capture personality in non-Western contexts.

Issues related to the use of psychological tests in non-Western contexts

It is probably no surprise that intelligence, memory and critical thinking are influenced by the cultural context in which people live. Humans faced different challenges around the world in order to survive and this was particularly true before the modern era. Industrialization and the development of modern technology resulted to an increase in the need for people with specialized education and this continues up to the present time thus resulting to the adjustment of most of the education system in world in order to meet this need. Students now have to learn, understand and memorize, whether it be in mathematics, languages, or social studies. Students need to learn how to organize, assess and memorize information, and how to retrieve the appropriate knowledge from their memory whenever they need it.

Nevertheless, it is important for psychologists to recognize that this type of education is a component of one's culture and that culture varies in different countries and regions. Standardized psychological tests developed by Western psychologists are done so within a specific cultural context and to automatically apply them directly to non-Western samples can be misleading and problematic. When western researchers have used

standardized tests of cognitive abilities to participants from non-Western countries, almost all the time, the said participants performed poorly in many of the tests (Cole & Scribner, 1974). This clearly raises ethical and social implications since it can lead to people being judged as 'less able' or of 'low intelligence', not because they really are but because of the nature of the standardized test which they were asked to take. The 'local' validity of standardized tests, taking into consideration language and culture specific behavior, beliefs and attitudes, needs to be established first before psychologists use these tests in non-Western cultures.

The effect of culture on the assessment of cognitive abilities is clearly demonstrated by Cole's and Scribner's (1974) study, which investigated memory strategies in different cultures. The researchers compared the recall of a series of words for elementary school children from the USA and the Kpelle people of rural Liberia in West Africa. They understood that they could not use exactly the same list of words with the Liberian and American participants and in order to come up with a 'culture fair' test they observed the everyday cognitive activities of the Kpelle tribe in order to develop a memory tests that consisted of culturally relevant tasks. They made sure that the words used in their memory tests were familiar to the participants and collaborated with local college-educated people who took the role experimenters in the study. In spite of these precautions, their findings showed profound cultural differences in the way the Kpelle people remembered information and solved problems compared to the American participants.

In their first study the researchers asked Kpelle children from different age groups to recall as many items as possible from four categories: Utensils, Clothes, Tools and Vegetables. When researchers analyzed the data, they found that illiterate (e.g. non-schooled) children did not use strategies such as 'chunking'. Chunking is the grouping bits of information into larger units, to help them remember. The Kpelle did apply the mnemonic strategy of mental rehearsal, because the position of a word in the word list did not have an effect on the rate of recall. In their second study, the researchers changed the recall task so that the objects were now presented in a meaningful way in the form of a narrative story. They found that the illiterate (non-schooled) children recalled the objects easily and actually chunked them according to the roles they played in the story. This is of great significance to psychologists, as it highlights the importance of recognizing how culture influences performance on standardized tests of cognitive abilities. The example from Liberia suggests that it is vital for tests to be contextualized and made understandable to those who are taking the tests.

Further cross-cultural research comes from Kagan, Klein, Finley, Rogoff and Nolan (1978) who found that the memory abilities of Mayan children from Guatemala were significantly lower than their American counterparts based on standardized tests. However, Rogoff and Waddell (1982) decided to re-examine these findings in their study of cognitive abilities of Mayan children from Guatemala and American children from Salt Lake City. They aimed to show how placing a test of cognitive ability into the appropriate context could enable better performance. In their study each child watched as a local experimenter select 20 miniature objects from

a set of 80 objects and placed them in the 3D model diorama. The objects that were put in the diorama included animals, cars, people and furniture. The 20 objects were then returned to the group of 60 others on the table. The children were then asked to reconstruct the full scene they had been shown. They found that under these conditions, the memory performance of the Mayan children was slightly superior to their American counterparts. They concluded that children of any culture are skilled at remembering if information is presented in a meaningful context.

Critical thinking is another construct where cultural differences are present. Lun, Fischer and Ward (2010) analyzed differences in critical thinking between Asian and European students in New Zealand. They found that the European students performed better than the Asian students on the Halpern Critical Thinking Assessment and the Watson Glazer Critical Thinking Appraisal. The researchers concluded that these differences are due to Asian students relying more on *dialectical thinking* than their Western counterparts, due to differences in the way they process information. According to Nisbett, Peng, Choi, and Norenzayan (2001) Asians tend to 'perceive more changes, are more tolerant to contradictions and see things as more interrelated than westerners', they also concluded that westerners are more likely to use formal logical rules in reasoning and Asians use more experience based intuitive reasoning (Lun et al., 2010, p.605).

The above studies suggest that memory, intelligence and critical thinking have both etic and emic aspects. It is therefore vital for psychologists to understand this when administering standardized tests. Past researches demonstrates that people solve problems, learn and remember things in ways that are relevant to their own everyday lives, which do not always fit in with the way the concept is articulated in Western psychological tests.

Overcoming issues related to the use of psychological tests in non-Western contexts

When it comes to the use of standardized tests outside the context in which the test was developed, the fundamental issue is *equivalence*. If psychologists are to adapt or use standardized tests in different cultures fairly and accurately they need to ensure that translation equivalence, conceptual equivalence and metrical equivalence have been fulfilled (Brislin, 1993). It is also important to develop local norms for these standardized tests and to promote the development of locally made tests that are relevant to the test takers cultural background.

The first step in translation equivalence is to ensure that the concepts being measured in the items of the test are easily expressed in the language of the non-Western culture. If the test items do not translate well it suggests that the items are not relevant to that particular cultural context. A common method to ensure translation equivalence is *back-translation*. For this process to be carried out the original test items are translated from English into the target language by a bilingual person and a second bilingual person then translates it from the target language back into English. If the two English versions are equivalent it is assumed that the target language version is adequate.

Conceptual equivalence begins with the assumption that there will be different aspects of a concept, which serve the same purpose in different cultures. In order to ensure conceptual equivalence the etic and emic aspects of the constructs needs to be identified. For example, for the etic of 'intelligence' the emic might be 'solving problems the exact form of which have not been seen before' while the Western emic might include 'mental quickness' in the USA and Western Europe while it is 'slow careful and deliberate thought' for the Baganda people of Uganda (Wober, 1974). The different emic here are conceptually equivalent because they all form part of the definition of intelligence used in the two cultures. When it comes to standardized testing, in order to ensure conceptual equivalence both the emic and etic aspects of the construct need to be assessed for the cultural context for which the standardized test was created and also in the target culture in which the test will be used.

Metric equivalence focuses on the analysis of the same concept across cultures, based on the assumption that the same scale (after proper translation procedures have been carried out) can be used to measure it. For example, after back-translation, an IQ test produces a score of 120 for an American woman and a woman in the Philippines. The assumption is that the intelligence scale is measuring exactly the same concept (intelligence) in the two countries and that a score in one country can be directly compared with a score in another.

Of the three equivalences outlined above, conceptual equivalence is the most important criterion for standardized testing because of its theoretical impact. Another way to overcome these issues is to have local test constructors make their own standardized tests but then this approach limits the ability to compare the results with international norms. The issues outlined above raise the question of whether it is even possible to have 'culture fair' standardized tests. According to Frijda and Jahoda (1966) it is possible to have a culture-fair test only if it fulfills the following criteria: (i) if it has a set of items that are equally unfamiliar to all possible persons in all possible cultures, so that everyone would have an equal chance of passing (or failing) the items; or (ii) if it comprises of multiple sets of items, modified for use in each culture to ensure that each version of the test would contain the same degree of familiarity. This would give members of each culture the same chance of being successful with their respective versions of the test. Unfortunately, criteria (i) is virtually impossible and criteria (ii) is possible in theory but in practice very difficult to construct.

Conclusion

It is clear that there are real difficulties faced when using standardized tests developed in Western countries in non-Western cultures. Psychologists need to be aware of them and whenever Western standardized psychological tests are used, equivalence needs to be established first via the rigorous process of test validation. Once the tests are made as equivalent as possible it is crucial that local norms are established so that the results of one group of people are not unfairly compared to a different group that is from a different cultural setting. It is also vitally important to

create more culturally relevant tests in local contexts, which will be of practical use to psychologists in their assessments.

References

- Brislin, R.W. (1993). *Understanding culture's influence on behavior*. Fort Worth, TX: Harcourt Brace Jovanovich.
- Cattell, P. (1973). *Measuring intelligence with culture fair tests*. Metro Manila: Philippine Psychological Corporation.
- Cole, M., & Scribner, S. (1974). *Culture and thought: A psychological introduction*. New York, NY: Wiley.
- Costa, P.T., & McCrae, R.R. (1992). *Revised NEO Personality Inventory and NEO Five-Factor Inventory Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Digman, J. M., & Inouye, J. (1986). Further specification of the five robust factors of personality. *Journal of Personality and Social Psychology*, 50, 116-123.
- Flanagan, J. C. (1960). *Flanagan industrial tests*. Metro Manila: Science Research Associates.
- Frijda, N., & Jahoda, G. (1966). On the scope and methods of cross-cultural research. *International Journal of Psychology*, 1, 109-127.
- Funder, D. (2010). *The Personality Puzzle*. New York, NY: WW Norton & Company.
- Glick, J. (1975). Cognitive development in cross-cultural perspective. In F.D. Horowitz (Ed.), *Review of child development research* (pp. 595-654). Chicago, IL: University of Chicago Press.
- Guthrie, G. M., & Bennett, A. B., Jr. (1971). Cultural differences in implicit personality theory. *International Journal of Psychology*, 6, 305-312.
- Herskovits, M.J. (1948). *Man and his Works: The Science of Cultural Anthropology*. New York NY: Knopf.
- Hofstede, G. (2001) *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Thousand Oaks, CA: Sage Publications.
- Kagan, J., Klein, R. E., Finley, G. E., Rogoff, B., & Nolan, E. (1979). A cross-cultural study of cognitive development. *Monographs of the Society for Research in Child Development*, 44, 5.
- Katigbak, M. S., Church, A. T., Guanzon-Lapeña, M. A., Carlota, A. J., & del Pilar, G. H. (2002). Are indigenous personality dimensions culture specific? Philippine inventories and the five-factor model. *Journal of Personality and Social Psychology*, 82, 89-101.
- Lun, V. M. C., Fischer, R., & Ward, C. (2010). Exploring cultural differences in critical thinking: Is it about my thinking style or the language I speak? *Learning and Individual Differences*, 20, 604-616.
- Matsumoto, D. (2003). Cross-cultural research. In S. Davis (Ed.), *The handbook of research methods in experimental psychology* (pp. 189-208). Oxford, UK: Blackwell.
- McCrae, R. R. (2004). Human nature and culture: A trait perspective. *Journal of Research in Personality*, 38, 3-14.

- McCrae, R. R., & Terracciano, A. (2005). Universal features of personality trait terms from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology* 88, 29-44.
- Nisbett, R.E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, 108, 291-310.
- Otis, A. S., & Lennon, R. T. (1957). *Otis-Lennon mental ability test*. New York, NY: Harcourt Brace Jovanovich.
- Pike, K. L. (1954). *Language in relation to a unified theory of the structure of human behavior* (2nd. ed.). The Hague: Mouton.
- Popham, W.J. (1999). Why standardized tests don't measure educational quality. *Educational Leadership*, 56, 8-15.
- Raven, J. (1936). *Ravens advanced progressive matrices*. Metro Manila: Philippine Psychological Corporation.
- Rogoff, B., & Waddell, K. J. (1982). Memory for information organized in a scene by children from two cultures. *Child Development*, 53, 1224-1228.
- Segall, M. H., Dasen, P. R., Berry, J. W., & Poortinga, Y. H. (1999). *Human behavior in global perspective* (2nd ed.). Boston, MA: Allyn & Bacon.
- Shaules, J. (2007). *Deep culture: The hidden challenges of global living*. Clevedon, UK: Multilingual Matters.
- Smith, P. B., & Bond, M. H. (1998). *Social psychology across cultures* (2nd ed.). Hemel Hempstead, UK: Prentice Hall.
- Triandis, H. C. (1990). Cross-cultural studies of individualism and collectivism. In J. Berman (Ed.), *Nebraska Symposium on Motivation, 1989* (pp. 41 - 133). Lincoln, NE: University of Press of Nebraska.
- Wober, M. (1974). Towards an understanding of the Kiganda concept of intelligence. In J.W. Berry & P.R. Dasen (Eds.), *Culture and cognition*. London, UK: Methuen.

Correspondence may be sent to:
MarshallDy@ismanila.org