Running head: Testing and assessment in an international context

Chapter 29

Testing and assessment in an international context: cross- and multi-cultural issues

Jérôme Rossier

University of Lausanne, Switzerland

M. Eduarda Duarte

University of Lisbon, Portugal

Abstract

Globalisation, increase of migration flows, and the concurrent worldwide competitiveness impose rethinking of testing and assessment procedures and practices in an international and multicultural context. This chapter reviews the methodological and practical implications for psychological assessment in the field of career guidance. The methodological implications are numerous and several aspects have to be considered, such as cross-cultural equivalence or construct, method, and item bias. Moreover, the construct of culture by itself is difficult to define and difficult to measure. In order to provide non-discriminatory assessment, counsellors should develop their clinical cross-cultural competencies, develop more specific intervention strategies, and respect cultural differences. Several suggestions are given concerning translation and adaptation of psychological instruments, developing culture specific measures, and the use of these instruments. More research in this field should use mixed methods, multi-centric designs, and consider emic and etic psychological variables. A multidisciplinary approach might also allow identifying culture specific and ecological meaningful constructs. Non-discriminatory assessment implies considering the influence and interaction of personal characteristics and environmental factors.

Testing and Assessment in an International Context: Cross-and Multi-Cultural Issues

Globalisation, increase of migration flows, and the concurrent worldwide competitiveness impose a shift on intercultural research, towards an integrative background to both common and regional competencies to achieve added value and usefulness of testing and assessment techniques. Whenever a test is translated and adapted for use in another language or culture, it is mandatory to know that the process begins always with evaluation: the first issue concerns the conceptual definition and the context of its operationalisation, which means the identification of the relevant contents of cultural knowledge. In this case, the ingredients of knowledge mean the understanding of how culture is expressed through beliefs and values, behavioural expressions, symbols and habits, but also mean a balance between cognitive knowledge and attributes of *good judgement* to deal with the culture or sub-culture variables. The close connection between these two aspects can lead to a deep awareness of construct validation research within each population for which translation or adaptation occurs.

As Hambleton (2005) pointed out, a distinction has to be made between test adaptation and test translation. In his view, the term "adaptation" is broader and more reflective of what should happen in practice when preparing a test that is constructed in one language and culture for use in a second language and culture. Test adaptation includes all the steps involved in developing a measure to assess a specific construct in a different language and culture, taking into account the specificities of that language or culture (Arbona, 2014). Such an adaptation may imply to develop new items, consider new subscale, or to redefine the construct. Translators are trying to find concepts, words and expressions that are culturally, psychologically, and linguistically equivalent in a second language and culture, and so clearly the task goes well beyond simply preparing a literal translation of the test content (Osborn, 2012). In short, test adaptation does not run in straight lines: evaluative

information on culture and context is much more complex than creating guidelines for cross-cultural normative assessment.

The classical discussion about "culture-free" tests (Cattell, 1940), "culture-fair" tests (Cattell & Cattell, 1963), or "culture-reduced" tests (Jensen, 1980)—an important debate that took place during almost 50 years—belongs to the past (Duarte, 2005). Cultural issues need to be understood as meanings and practices that have an important role and mediate the impact of ecological suitability. It seems interesting to refer to the origin of the word *ecology*: from the Greek etymon *oikos* (house) and *logos* (order as intrinsic rationality). Ecology might be translated by "put one's house in order". Coincidently, the word *economy* has the same first etymon *oikos*, and *némó* (that means distribute the spaces). Then, in a broader sense, ecology and economy, having commonalities pointing to the same sense: a way in which things are placed in relation to one another—that means appropriateness or suitability. Indeed, the challenge today is to have non-discriminatory assessment procedures that are appropriate in our multi-cultural societies.

Nowadays, test adaptation, even considering the "free", the "fair", the "reduced" or other added words to culture, does not address the presumed equivalence of constructs in a different context that differs from the environment in which the original test was developed. This implies three main issues to assess behaviour in a particular culture: first, test development should be based on situation sampling, through the definition of the relevant and observable aspects of a particular construct; second, test development should be based on function sampling, through the refinement of test items in terms of how they could be operationalised within a specific cultural context; and third, test development should be based on the identification of differential variables and context information (e.g., patterns of cultural or subcultural rewards).

Cultural bias or a poor understanding of how culture influences the process of translation or adaptation can generate distortions on the assumption of the differential paradigm, centred on the rationale of individual differences to determine the experimental design and decisions about psychometric procedures. From the moment that theoretical concepts are translated into and assessed as measurement dimensions, a variety of procedures must be examined: evaluative information on culture (seen as a construction) and context; stipulation of relevant and observable aspects of the construct; items phrasing (operations and content choices); differential variables and context information.

Creating ways of collecting information that is pertinent regarding cross and multi-cultural issues also comprises methodological questions. These issues are examined as well as the outcomes of testing and assessment in an international context. During the last decades, a recurring criticism among many career counsellors is that standardised tests are culturally biased. The concept of culture as a socially constructed phenomenon, and questions like the sampling of behaviour across cultures will be discussed. This chapter also presents the difficulties and challenges of translating and adapting psychological measurements based on the experience of processes of adaptations of assessment devices. The debate of non-discriminatory assessment implies the development of culture specific measurements or of instruments that are simultaneously created in several cultures. The Work Importance Study (Super & Šverko, 1995), the development of the Career Adapt-Abilities Scale (Savickas & Porfeli, 2012) or the current international project about decent work (Duffy & Blustein, January 2017, personal communication) are examples of construction of instruments adapted to various ecological contexts.

Globalization and migratory flows induce a cultural diversification of societies that push vocational psychologists to strengthen some specific lines of research: promote cross-cultural studies analysing the relationships between psychological measurements and external

criteria; or ethnographic multi-method methodologies to conceive new culturally specific theories and quantitative and qualitative measurement tools in the field of career guidance. The chapter concludes with practical implications for the use of measurement procedures and instruments with culturally diverse populations. Research on testing and assessment in an international context will have to also consider interdisciplinary, contextual, functional, and processual aspects, the interaction between the different life domains, new theories in our field, and adopt a holistic perspective.

Methodological implications

The goal of offering a more comprehensive approach to assessment in order to describe better the strengths of diverse populations has several methodological implications. The culture-fair perspective implies, for example, the simultaneous development in a variety of cultures of measurements based on a model that has the potential to be transposed into different cultural settings. The Five-Factor Model or the Alternative Five-Factor Model of personality are examples of such models that replicated well across many countries (McCrae, 2017; Rossier et al., 2016) even if, in some cases, minor variations could be observed in some specific cultures (Rossier, Ouedraogo, & Dahourou, 2017). Interestingly, translation in some specific languages can be quite challenging (Rossier, Ouedraogo, Dahourou, Verardi, & Meyer de Stadelhofen, 2013). Recently, a combined emic-etic approach was used to simultaneously develop, in different ethnic groups, the South African Personality Inventory (SAPI; Fetvadjiev, Meiring, van de Vijver, Nel, & Hill, 2015). This study allowed the identification of personality dimensions that seem to be universal and other more culture-specific, and the frequency of the use of personality descriptors seemed to vary across cultural groups. However, all theories do not have this potential to be easily transposed. In particular, where models are culturally founded, such as value systems or career aspirations and expectations (Metz, Fouad, & Ihle-Helledy, 2009), culture specific measurements should

be developed. "[Each] language and group has unique value terms, but all can be located on the circular continuum and subsumed under one of the basic values" (Schwartz, 2017, p. 129). Furthermore, several models might be transposed to some cultural settings but are not universal. This seems to be the case for Holland's vocational model, for example (Armstrong, Hubert, & Rounds, 2003; Ryan, Tracey, & Rounds, 1996). In this context, two concepts are of prime importance to assess the appropriateness of measurement instruments across cultures: first, the notion of equivalence, and, secondly, the notion of bias (Van de Vijver & Leung, 2011).

Any cross-cultural or cross-national studies can be considered as quasi-experimental because the different groups cannot be distinguished according to only one independent variable, as for example language, because they are not similar in all other respects. In fact, in the case of cross-cultural or cross-national studies, existing groups are compared (Van de Vijver & Matsumoto, 2011). In this type of study, the control on the independent variables is much weaker and should imply a description of what distinguishes two cultures, which somehow seems difficult. One implication of this difficulty is that the reasons of observed differences between groups of subjects from different cultures are difficult to identify. "Culture is too global a concept to be meaningful as an explanatory variable, however, and should be replaced by its constituents" (Van de Vijver & Leung, 1997, p. 3). These constituents called context variables by Van de Vijver and Leung can be person-related, such as age or gender, culture-related, or nation-related like gross national product. The idea is roughly to identify variables that might account for cross-cultural score differences and that might also be used as indicators of external validity. One example of such an approach can be seen in the research conducted by Ryan and colleagues (1996), who studied Holland's structure of vocational interests across ethnicity, gender, and socioeconomic status.

The concept of equivalence is closely associated with testing and assessment in a cross-cultural context and concerns the comparability of scores obtained in different cultural settings. Several levels of equivalence across cultures are considered but the labels used vary in the cross-cultural literature, inducing some confusion. Most of the time, three levels of psychometric equivalence are considered. The first level is a level, that may be called *configural equivalence*, and that imply to verify if a model can be relevant for different groups. This level is called structural or functional equivalence by van de Vivjer and Leung (2011). The idea is thus to vey if a factor structure is similar across groups, using multigroup analysis. In this case we can speak of a non-metric equivalence, because is assesses only overall configuration of the structure of the studied construct. The second level is commonly called *metric equivalence* and imply to verify that the factor loadings are equal across the groups (Byrne, 2016). In case of equivalence of the loadings across groups, we may speak of partial measurement equivalence. In this case the metric of the scales should be equivalent across groups. Finally, the third level is commonly called *scalar equivalence* and imply to verify that intercepts are equivalent across groups. According to Meredith's (1993) nomenclature, metric equivalence would fall into the *weak* and scalar equivalence into the *strong* metric invariance category. The level of equivalence is certainly dependent on the construct measured, on the characteristics of a measurement instrument, but also on the cultural distance between the studied groups. The three levels of equivalence have to be considered with great caution. If configurational equivalence does not imply metric or scalar equivalence, scalar equivalence implies the two less demanding equivalence levels. Across cultures, scalar equivalence is usually difficult to reach, but would be necessary for testing mean differences (Byrne, 2016). For this reason, several methodologists suggested alternative less strict method to assess measurement invariance, especially when the goal is to test latent mean differences (Millsap, 2011; Raykov, Marcoulides, & Li, 2012). Other authors have

proposed other more liberal alternatives, such as a Bayesian approach to assess approximate

measurement invariance (e.g., Zercher, Schmidt, Cieciuch, & Davidov, 2015).

Bias might affect all steps of research aimed at developing measurement instruments.

Measurement instruments are themselves affected by several types of bias. Bias can affect

theoretical constructs, research procedures, or data analysis. Using a cognitive ability test

with populations that do not benefit from similar school systems might lead to differences in

stimulus familiarity that can explain the cross-cultural score differences. Van de Vijver and

Leung (2011) distinguished three types of bias: *construct bias*, *method bias*, and *item bias*.

Construct bias concerns, for example, incomplete overlap of the definition of the construct

across cultures. Construct under-representation or, in other words, a poor sampling of aspects

relevant to a specific construct, might lead to such a construct bias. Another source of

construct bias is the transposition of Western constructs to non-Western cultures where these

constructs may be less relevant or have to be defined differently. For example, transposing an

interest inventory to a country where career options are limited, might lead to such a bias. As

interests develop in interaction with their environment, the absence in a given environment of

the possibility of exercising a certain number of professional activities could have an impact

on the structure of professional interests. However, some instrument can be remarkably stable

across two very different cultures, as the *Personal Globe Inventory* or the *Career Decision-

Making Difficulties Questionnaire* (Atitsogbe, Moumoula, Rochat, & Rossier, submitted).

Even if construct bias is avoided, an instrument can still be affected by method bias, such as

differences in social desirability or differences in stimulus or response procedure familiarity.

This bias is closely associated with the characteristics of an instrument. For example, an

interest inventory might ask the respondents to rate several professions that are not equally

familiar for people of different cultures. Moreover, response style, as socially desirable

responding or acquiescent response bias, are affected by culture and might be an aspect of

cross-cultural communication style (Johnson, Shavitt, & Holbrook, 2011). All these sources of method bias have an impact on the mean levels measured on the assessed dimensions, and they cannot be attributed to individuals. Finally, item bias refers to differential item functioning and can be due, for example, to poor translation or inadequate item content.

Now, considering the impact of bias on equivalence, no equivalence can be attained in case of construct bias. However, in the case of method or item bias, structural equivalence and even, in some cases, measurement unit equivalence might be reached, as long as the bias affects all items uniformly (Van de Vijver & Leung, 2011). For this reason, it is very important to identify these biasing effects when using a measurement instrument in an international context.

Most commonly used assessment instruments in the field of career guidance have been translated into numerous languages. For example, the *Strong Interest Inventory* and Holland's *Vocational Preference Inventory* and *Self-Directed Search* have been translated into more than 15 languages (Rounds & Tracey, 1996). However, several studies suggested that Holland's hexagonal model does not replicate well across cultures (e.g., du Toit & De Bruin, 2002; Long & Tracey, 2006; Rounds & Tracey, 1996) suggesting that indigenous interest structures should be considered (Einarsdottir, Eyjolfsdottir, & Rounds, 2013). A more recently developed spherical model of interests and its measurement instrument called the *Personal Globe Inventory* (PGI; Tracey, 2002) seem to replicate quite well across cultures and were validated in various socio-cultural contexts, such as in Japan, Serbia, or Bulgaria (Hedrih, Stošić, Simić, & Ilieva, 2016; Long, Watanabe, & Tracey, 2006). A study having investigated the measurement equivalence of the PGI across Switzerland and Burkina Faso has shown that the PGI did reach configural and metric equivalence but not scalar equivalence, suggesting that culture specific norms should be considered (Atitsogbe et al., submitted). These studies indicate that some interest models and inventories appear to be

quite sensitive to the cultural context whereas the spherical model seems to be more robust. It has to be noted that many locally developed interests inventory assess culture-specific dimensions (Rossier & Fiori, in press).

The study of the measurement equivalence of Schwartz's structural model of values is another illustration of this type of research about the cross-cultural generalisation of a measurement instrument. Schwartz (1992) defined a model of values and developed the *Schwartz Value Survey* on the basis of empirical cross-cultural studies (e.g., Schwartz & Bilsky, 1990). This model defines 10 values that can be represented in a bidimensional space. This structure was found to be very stable across cultures even if some consistent deviations that might represent some potential culture-specific characteristics have been observed (Schwartz & Sagiv, 1995). More recently Heim and colleagues (2017) studied values of Chinese, Russian, and German students and observed that the Portrait Value Questionnaire (PVQ-21) did reach configural, metric, and partial scalar invariance. However, their expectations in terms of cross-cultural differences were only partially met, suggesting that the relative importance of values could have changed due to the recent economic and cultural developments. For Schwartz (2017) the language and group terms have unique value:

> but all can be located on the circular continuum and subsumed under one of the basic values. Individuals and groups differ in the priority they ascribe to particular values, but not in the content and structure of the values they recognize. (p. 129)

Taking into account these concepts of bias and of equivalence has several practical implications for researchers developing measurement instruments. For example, when developing an instrument in several cultures simultaneously, attention should be paid both to the sampling of cultures and the sampling of subjects. In order to have a diverse set of cultures and to maximise the chance of identifying cultural differences, it is desirable to

select cultures as different as possible and simultaneously to maximise the comparability of the subjects across cultures. Moreover, several statistical techniques have been developed to assess equivalence of tests or items across culture. An item bias analysis verifies that individuals from different groups supposed to have *equal standing* on a particular construct have similar scores on items intended to assess this construct. This can be done by using item response theory or by analysing the measurement equivalence, using exploratory factor analyses or structural equation modelling (Leong, Bartram, Cheung, Geisinger, & Iliescu, 2016; Matsumoto & Van de Vijver, 2011).

Challenges of non-discriminatory assessment

The challenges of non-discriminatory assessment consist of identifying strengths and weaknesses of individuals being assessed without any influence due to their belonging to a specific culture or minority group. For this purpose, it might be very useful to take into account how individuals relate and interact with various systems existing in their own proximal world. However, if scalar equivalence could be attained, these precautions should cease to be necessary, but scalar equivalence is only very rarely observed. For this reason, the potentials and difficulties of culturally diverse clients might not be correctly assessed using traditional standardised measures (Schwabe, von Davier, & Chalhoub-Deville, 2016). The challenges of non-discriminatory assessment become an even more acute topic with the important increase nowadays of migratory flows (OECD, 2017) that implies that some of these people will need career guidance and counselling to sustain their social integration.

Each culture or minority group can have specificities that should be taken into account when using a measurement instrument. The principle of equity implies the recognition of this diversity and career counsellors should select measurement instruments that are adapted to these groups. Indeed, the challenge for equity in assessment implies that the measures are accurate and that the interpretations and decisions made on the bases of these measures do

not discriminate any of these groups. However, the task of a researcher designing an assessment procedure is made particularly difficult because societies themselves do not generally respect the principle of equity. A review of the principles generally recommended for cross-cultural assessment concern three aspects: competencies, intervention strategies, and respect for cultural differences (Gopaul-McNicol & Armour-Thomas, 2002). An example of such recommendations can be found in the "Guidelines for providers of psychological services to ethnic, linguistic, and culturally diverse populations" (American Psychological Association [APA], 1990). For adapting tests, the guidelines of the International Test Commission (2010) should be considered. Concerning competencies in the field of cross-cultural assessment, it is necessary that career counsellors understand the limits of the assessments they use with culturally diverse groups. When interpreting the results, they should take into account cultural factors that might impact clients' scores. When using measurement instruments with culturally diverse groups, it is of course crucial to respect cultural differences and to adapt intervention strategies or the measurement procedures to the culture of the client. Indeed, cultural practices may have an influence on an individual's performance or response to specific assessment tools.

Non-discriminatory assessment implies that career counsellors take the influence and interaction of personal characteristics (e.g., interests, values, abilities) and environmental factors (e.g., social status) into account (Collins & Arthur, 2010). Indeed, all these factors might impact the career choice and the career path of an individual. For example, several aspects of a client's environment, like his/her parents', sisters', or brothers' professions, may have a substantial influence on their vocational interests. For this reason, it is important to consider that such factors may contribute to the choice of a particular job. However, in a meta-analysis that investigated the relation between culture and vocational choice variables, Fouad and Byars-Winston (2005) observed that cultural differences do not greatly affect

career aspirations but that these differences might affect the perception of career prospects, which might be in agreement with the socio-political context within which these minority groups are living. This study suggests that career counsellors should pay special attention to these perceptions that seem to be context or culture specific.

In order to promote non-discriminatory assessment or to reduce inequity, several recommendations might be given to career counsellors when working with clients in an intercultural setting. The first measure might simply be to ensure that everyone has equal access to career guidance programs. In order to reduce inequity in assessment, career counsellors should pay attention to the cross-cultural validity of the instruments they use, be sure that appropriate norms are available, use the appropriate language version, spend more time on exploration, use different types of instruments, adapt the assessment procedure, and compare the results with other information obtained using a clinical approach.

Concerning standardised measurement instruments, the cross-cultural validity should be more systematically assessed or confirmed. It is also crucial to use norms adapted to the culture or the minority group in order to ensure the fairness and accuracy of the interpretation of the client's test results (e.g., Rossier, 2005). Culture-specific norms should correct for social inequality and unequal opportunities in societies. It is, of course, important that clients are assessed in their dominant language. The effect of the cultural environment might be especially strong if the dominant language of clients is not the usual language spoken in their environment (Gopaul-McNicol & Armour-Thomas, 2002). If it is not possible to assess the counselees in their dominant language, it is necessary to determine the level of proficiency in the language used by the proposed assessment tools. A non-verbal test can be an option in some cases, even if culture-specific norms are also necessary. To assess abilities, the *Wechsler Nonverbal Scale of Ability* (Wechsler & Naglieri, 2006) or the *Naglieri Nonverbal Ability Test* (Naglieri, 2016) are often used.

For interests, values, or to screen contextual and personal strengths and weaknesses or resources and vulnerabilities, qualitative assessment can be very useful with diverse populations (Goldman, 1992). For interest, card sorting tasks are also interesting tools and allow one to understand a counselee's criteria of choices. They stimulate vocational exploration and can be used for people with low verbal skills or who have a poor representation of vocations. Moreover, interest exploration should also expose the client to a variety of careers in a very concrete way (for example, visiting job sites or participating in training periods might be useful in such cases). This might be especially effective for clients that were never directly exposed to some specific professions. Roughly, the main idea is to minimise the chance that clients' choices are not based on an exploration and evaluation of all effective possibilities offered by their environment. This risk of a foreclosed choice is especially important for migrants and their families who might be less familiar with the vocational and educational system of their host country.

Inventories can also be used by adopting a more clinical approach; items that would not be understood could be rephrased by the counsellors in order to make them fit with the client's cultural realm, educational or social experiences:

> It is therefore critical for 'culture fair' vocational assessors/examiners to be aware of the questions on each test that may present some difficulty for culturally diverse children [or clients] and to assist them with each of these questions by engaging in an item equivalency type of approach. (Gopaul-McNicol & Armour-Thomas, 2002, p. 117)

Finally, when assessing migrants, it could be advisable to assess the adaptation process, by assessing counselees career development using repeated measures of competences, values, projects, barriers, etc.

In some cases, it might also be interesting to compare results of traditional psychometric tests with curriculum-based assessments and portfolios. Curriculum-based assessments consist of assessing an individual several times during a learning process or to analyse the abilities based on real-life achievements, and a portfolio is a self-evaluative tool asking the subject to list all his/her competencies and to document them. For example, leadership or organisational competencies might be documented with voluntary activities. Another concrete precaution, in the case of measurement instruments with time limits, as with intelligence measures, is to let the client continue after the time limit in order not to have only the maximum performance but also an estimation of their potential without this time constrain (Gopaul-McNicol & Armour-Thomas, 2002). This seems particularly interesting with people who are emotional or unfamiliar with the assessment situation. This type of procedure is already proposed for some instrument like the *Wonderlic Personnel Test* (Wonderlic Inc., 1983).

Finally, career counsellors should go beyond standardised tests and use, for example, a clinical perspective. Several authors, have adopted and developed narrative and contextual approaches to assess clients from diverse cultures (Busacca & Rehfuss, 2017; Laher & Cockcroft, 2017). Yasui (2015) developed a process of clinical practice based on cultural experiences that foster cultural exchange. The goal of such a procedure is to obtain information about various aspects of a client's life, which might be affected by cultural factors in order to incorporate this information into the career counselling process. This kind of interview might help both the client and the counsellor to understand the impact of these factors on career decisions. Moreover, career counsellors have to be familiar with the culture of their clients. All these suggestions do not warrant controlling for all bias in cross-cultural assessment but should allow reducing inequity.

It would of course be much easier to provide non-discriminatory assessment if assessment procedures were culturally equivalent. However, this would not solve all the problems and it seems very desirable in all cases to combine a standard evaluative approach with a multi-source and clinical approach. For this reason, the training of career practitioners should include multi-source career assessment combining quantitative, qualitative, and clinical evaluations. However, the endeavour of developing assessment tools usable with individuals from different cultures should be encouraged. Another way to increase the adequacy of assessment procedures would be to more systematically adapt versions of measurement instruments to fit diverse cultural realities (Leong & Brown, 1995). One advantage of creating culture specific instruments would be to avoid the risks of ethnocentrism (Marsella & Leong, 1995). Thus, combining the etic and the emic approaches seem to be a promising perspective. Moreover, more systematic research about constructs in the field of career guidance is needed.

Translating and adapting psychological instruments

When Alfred Binet (1857-1911) published his first intelligence scale in 1905, he certainly did not pay too much attention to ascertaining that the procedures are understood in identical ways in different cultural populations. He was far from the identification of cultural parameters that may affect the operation of the presumed universal psychological process (Kitayama, 2002; Munroe, Munroe, & Whiting, 1981). In other words, cross-cultural methodological issues relevant to a rationale for developing international psychological devices were not a priority in the beginning of the 20th century. Since the creation of the first (in contemporary terms) test, the accountability of "imported" measurement devices, became, and still is, a major issue in psychology theory and practice.

At least five major reasons can be found for adapting tests: (a) it is cheaper and faster than constructing a new test; (b) when the purpose is cross-cultural, it is the most effective

way to produce an equivalent test, allowing comparing results across cultures; (c) lack of expertise for developing a new test; (d) sense of security, especially when the original test is well-known; and (e) fairness to examinees resulting from the presence of multiple language versions (Hambleton & Patsula, 1999). These five assumptions are still controversial; but, at least one more reason should be added: the recognition of the global prevalence of American models in psychological research, since the beginning of the 20th century until now and in particular its impact on both maximum performance and typical performance measures (Cronbach, 1990).

The advances and the recognition of the importance of inter-and intra-cultural differences in human behaviour in coping with environmental needs and pressures, led to a project initiated in 1992 by the International Test Commission (2010) aiming at the development of general guidelines for translating and adapting educational and psychological tests. Since then, steps for implementing the guidelines were taken, and some paths were opened leading to a more global approach to psychological measurement. Is it psychologists' way to respond to the effects of globalisation? Or is it an opportunity for taking new theories for the development of cross-cultural tests? There are good reasons to suspect that both questions bring to mind the same answer. In general, there are many differences among cultures, among regions, among countries, but there are also commonalities and "the engagement in dialogue about international perspectives on and comparative features of educational and vocational guidance around the globe provide a comprehensive understanding of the issues faced by scholars and specialists concerned with the internationalisation of educational and vocational guidance" (Savickas, Van Esbroeck, & Herr, 2005, p. 84). Many practitioners and scholars around the world have expressed the importance that colleagues from all over word participate in the development of indigenous testing and assessment procedures, tools, and career development models (e.g., Leong &

Pearce, 2014). In summary, test results in conjunction with the interpretation of cultural values can be used to develop theoretical and empirical studies for the purpose of being useful to individuals who need career psychological assistance.

Testing and assessment in an international context relies on methodological requirements for cross-cultural equivalence and cross-cultural adequacy. Items and constructs should replicate well but also have the same relevance from one context to the other. In this sense, it seems desirable to propose a comprehensive framework for the implementation of the adaptation process—a process that has to take into account the specificities of the context towards which the adaptation is done. Some general steps to do *the job* are discussed based on the experience in translating/adapting psychological measures, such as *The Adult Career Concerns Inventory* (ACCI) (Super, Thompson, & Lindeman, 1985) or the *Career Adapt-abilities Scale* (Savickas & Porfeli, 2012), translated and adapted into several languages, but also in the literature connected with translating and adapting psychological tests (e.g., Hambleton, 2001; Hambleton, Merenda & Spielberger, 2005; International Test Commission, 2010; Oakland, 2004).

Step 1. Translating an observation device into another language is more than lexical transposition. The difficulties and challenges that the researcher encounters in preparing a test to be used in another language start from the moment of the decision to translate. If the classical translation procedures are used (Gjersing, Caplehorn, & Clausen, 2010; Warner & Campbell, 1970), there is a trap: linguistic equivalence is not a guarantee that items represent exactly the same construct dimensions. Thus, in some cases the content has to be adapted to insure construct validity. This aspect calls attention to the content of the components of the dimension intended to be measured, concerning behaviour and construct interpretation and meaning. It is a kind of exercise in order to make the bridge between the understanding of items phrasing in terms of operations and content choice (function sampling). Then, a first

point should be highlighted: the psychologist/translator should have an extended knowledge of theoretical literature and empirical studies related with the instrument he/she wants (or needs) to adapt.

Step 2. Draft translation of the test. The first attempt at translation should be accomplished using systematic methods and procedures (e.g., Duarte, 2005; Van de Vijver & Hambleton, 1996), including field-testing with the new respondents. This draft translation, as close as possible to the original version, should be administered in a non-standard way. Soliciting all kinds of opinions from the respondents (preferably, by psychologists well-versed in the theory undergirding the instrument) about the individual items, the interpretation of instructions and response alternatives, seems a good procedure for the launch of the adaptation process. However, cross-cultural applicability remains unproven. Opinions from natural groups (random sampling) are another point that should be taken into consideration; obviously, these natural groups have to belong to the target population for which the test was constructed. Butcher (1996), a researcher involved in studying personality assessment in intercultural contexts, and in particular, the use of the MMPI, refers to linguistic equivalence by using a seven-step procedure that includes translation, back translation, comparisons by bilinguals, field comparisons, adequacy with American norms, development of new norms with representative samples, and ongoing research to assess cultural validity. It seems that it was a very good intention, and scientifically very accurate and strong, but also shows unrealistic research possibilities.

Step 3. Amending. Adaptation is a sequential process of back and forward equivalence inspection and so forth. Linguistic meaning, cultural adaptation, and accurate technical information to reconstruct the translation are crucial issues at this moment in the job. The importance of field-testing to verify the acceptability in the target language is crucial. Another procedure is connected with construct validation, that implies among other the

assessment of structural validity. However, this methodology has been insufficient to make available unambiguous demonstrations (Strauss & Smith, 2009). It is thus of prime importance to assess construct validity within the target population after translation.

Step 4. Refinement of the adaptation process for the launching of the preliminary studies, and collection of empirical data. The support of experts in theoretical literature and empirical studies related to the instrument, in psychometrics, and experts well-versed in the test original language, as well as linguistic experts to guarantee superior standards of syntax and semantic is central to make the final revisions of the translation or adaptation process. After that, a pilot study should be carried out, using available expert participants to discuss the adapted version of the instrument, and also a sample of participants representative of the target population the test addresses. The researcher's work consists, again, in mapping convergent and divergent opinions related with the content, the format, and the response alternatives. The decision about ending the adaptation and launching preliminary studies depends on whether the researcher has found paths to answer the fundamental questions: does the construct exist, with the same components, in the adapted version? Are there differences in meaning between the two versions? Does the researcher have enough knowledge about the cross-cultural similarity of the construct? Of course, any researcher has no answers at this time of the process: evidence exists only with empirical studies. Statistical procedures after these preliminary studies may assume some importance; for example, comparisons with the results obtained with the original version, like reliability measures, and multiple-group factor analyses for checking construct equivalence.

Step 5. Administration to experimental groups. At this stage, probably the test is ready for the administration to a large group of participants representative of the target population. Linguistic procedures, elementary utilisation of psychometric apparatus, such as test and sub-test reliabilities, item-analysis, multigroup factor analysis can only provide an incomplete

demonstration of equivalence. However, this phase is crucial to verify the similarity, or equivalence, between the original and the adapted form of the test. Statistical analysis to determine construct equivalence between the two versions of the test is a procedure that tests whether the same dimensions underlie the scores. Until this precise moment of the translation or adaptation process, scalar equivalence demonstrations continue submerged (Van de Vijver, 2000), and only comparative empirical research can establish cross-cultural construct equivalence.

Step 6. Cross-cultural assessment. The cross-cultural study of structural equivalence is an important way of establishing the validity of the measures. Analyses of empirical evidence and data comparisons (Hambleton, 1993; Oakland, 2004) cover the major aspects related with situation sampling, function sampling, and ecological context. The administration of culture specific tests (Van de Vijver & Hambleton, 1996) has been accomplished by construct identification, measurement, and subsequent cross-cultural comparisons (Van de Vijver, 2015). For example, Atitsogbe and colleagues (submitted) assessed the cross-cultural invariance of both the *Personal Globe Inventory* (PGI) and the *Career Decision-making Difficulties Questionnaire* (CDDQ) in Burkina Faso and Switzerland and observed that both instruments reach configural and metric equivalence, but that only the CDDQ reached scalar equivalence. This suggested that culture-specific norms have to be considered for the PGI.

Extra step. And the job done. Adaptation methodology, in general, ends when empirical studies address the evidence of construct equivalence, as well as the absence of method and item bias. However, in cross-cultural comparisons it seems important to increase accuracy of data interpretation in order to enrich and develop new methods of practice in the career psychology assessment. Blustein (2006) admitted a healthy future for testing in the design and delivery of counselling services, and "believe[s] that a significant role for a

revitalised and culturally sensitive assessment process exists in expanding the reach and impact of our collective efforts" (p. 288). The role of testing and assessment in cross-cultural domains implies the design of new tools based on meaningful and culturally entrenched taxonomies, and the path is clear: not all constructs are universal, not all tests can be adapted.

When assessing a person, it is important to consider what is relevant in his/her cultural setting. From this point of view, it is possible to sketch a frame for psychologists working in assessment contexts with adapted forms of psychological instruments. The most important aspect is that we have to consider the difference between the general knowledge of the instrument(s), for example, about psychometric characteristics and metrological qualities, cross-national norms comparisons, meaning of interpretation results, and so on, and the information that describes an individual belonging to a specific group; putting things this way, tests results are "viewed from an explicit cultural framework in which the meaning of the items and the nature of the scores is embedded within the cultural understandings of the client's life space and worldview" (Blustein, 2006, p. 286). To do so, an overall conception of relevant issues connected with the knowledge of culture is needed. The case of Elaine presented by Savickas following the career construction theory (see, Savickas, 2004, pp. 60-68) could be used as an example. The theory of career construction "addresses how the career world is made through personal constructivism and social constructionism" (Savickas, 2004, p. 43). It is assumed that the theory has a universal conception. Elaine's problem is probably identical to other young college students around the western world, and all the steps done since the utilisation of *The Career Style Interview* (Savickas, 1989) are applicable with success in a great part of the world. The problem resides exactly in contextualising and "melding subjective and objective assessment data to comprehend and co-construct each individual's career path" (Hartung, 2005, p. 389). Only a *connoisseur* of the American culture is qualified to interpret and integrate Elaine's results of *The Career Style Interview*,

vocational personality, career adaptability, and life themes into a contextualised career counselling process. Trying to adapt measurements may also induce meaningful redefinition of our constructs and promote innovation.

A second issue related with adapted measures in cross-cultural studies involves the applicability of the measures to individuals of diverse backgrounds. This issue summarises the 68 Standards specifically relevant to the assessment of multicultural and diverse populations (Association for Assessment in Counseling [AAC], 2003). Cross-national evidence is definitely an added value to test validation, but more research focused on determining metrological qualities of the measures with equivalent cross-cultural samples of diverse groups is needed: only new empirical evidence can demonstrate the relevance of the utilisation of cross-national comparisons. The point is not to succumb to the temptation of ethnocentrism interpreted as "a belief that one's cultural ways are universally applicable in evaluating and judging human behaviour" (Baruth & Manning, 1992, p. 156). Cultural differences may affect the expectations and produce several sources of qualitative interpretation biases; interpretation is only practicable if the utilisation of the assessment data conforms to available normative data.

The process of development or adaptation into Portuguese (Duarte, 1995) of *The Adult Career Concerns Inventory* (ACCI) is presented. The ACCI assesses concerns with career developmental tasks in young and mature adults. The 61-item inventory yields scores for the career developmental stages of Exploration, Establishment, Maintenance, and Disengagement; and the sub-scale scores reflect developmental tasks within each stage (Crystallisation, Specification, Implementation, Stabilising, Consolidating, Advancing, Holding, Updating, Innovating, Decelerating, Retirement Planning, and Retirement Living). The 61-item instrument is a measure of the individual's career change status assessed by responses to five items. The participants indicate for each task how much concern they feel

currently, operationally assessed by responses to five items on a 5-point scale, from "no concern" to "great concern". The ACCI scores indicate the planfulness dimensions of Super's theoretical model of career ability (Super, 1990). Planfulness is the individual's skill to plan in a controlled way, a notion of self-esteem, and a strong awareness of the past for the preparation of the future. The reading level is established approximately at the eighth grade, although in specific situations it could be administered at lower levels of education.

The very first step for the Portuguese translation of the ACCI was as follows: a number of exercises and interviews were made in order to map the components of the dimension Planfulness, regarding behaviour and construct interpretation, and meaning of career concerns developmental tasks. The interviewees were psychologists, well versed in Super's theory, models and concepts (Step 1).

After that, a draft translation was presented in a non-standard way. The procedure was to solicit all kinds of opinions from psychologist respondents on all items, comprehension level, interpretation of the instructions, and response alternatives. Also, opinions from a random sample of young adults, college students, and employees between 23 and 65 years of age were collected. The first translation showed inappropriateness of some item content (Van de Vijver & Hambleton, 1996), particularly because of cultural adaptation. The decision not to use the occupational career fields of the original version was taken because of lack of correspondence with the Portuguese situation (Step 2).

Next, a field study was conducted in order to assess the acceptability of the translated items. This study was done with small samples, ranging from 60 to 175 participants, of male and female adult employees, of different age and occupational groups. Some problems of item formulation remained. This led to some adjustments to eliminate or re-adapt items leading to a different interpretation of the same statement (Step 3).

For the refinement of the adaptation process of the ACCI, the compatibility between the Planfulness dimensions (conceptual definition) and the operational definition measured by the ACCI (concerns with developmental tasks) was established by construct validity through correlational studies and factor analysis (Duarte, 1993). The results indicated that the ranking of concerns followed the theoretical ordering and supported the original interpretation of the scores. But, only a partial conclusion can be made with respect to demographic data, the ACCI represents a clear and theoretical picture of relationships between career stages and sub-stages concerns and age (Step 4).

Construct validation implies empirical demonstrations: the criteria of construct-related validation used with the original version of the ACCI were applied to the Portuguese version. The results obtained supported the appropriateness of the theoretical model, with respect to assessment of career concerns development (Duarte, 1993, 1995, 2005). The collected evidence supports the appropriateness of the theoretical model, with respect to assessment of career concerns development (Steps 5 and 6). However, the process of adaptation may also lead to new definitions of a construct and innovation.

Creating culture specific measures

The psychometric procedures, such as metric and scalar equivalence, or other statistical procedures can only provide incomplete demonstration of equivalence, and do not complete the process of construct validation, nor cross-cultural construct equivalence in absolute terms. The development of instruments that are simultaneously created in several cultures could be a way for the implementation research in cross-national domain, and it can also be a way to determine national differences in the assessed variables, and / or identify specific and common international, and /or regional patterns, achieving added value and usefulness of assessment techniques.

The Work Importance Study Project (WIS) (1979-1989), under the international coordination of Donald Super (Super & Šverko, 1995), constituted a very good example of ecological suitability in different ecological contexts (for more details see Duarte & Rossier, 2008). The International Career Adaptability project (ICAP) (Leong & Walsh, 2012) constitutes a good example of ecological suitability in different social contexts. The ICAP comprised a research project set up through an international team, and brought together researchers from 18 countries as follows: Australia, Belgium, Brazil, China, England, France, Germany, Hong-Kong, Iceland, Italy, Japan, Korea, Netherlands, Portugal, South Africa, Switzerland, Taiwan, and U.S.A. This international team did not start by creating a measure in one country and "then translate it for use in other countries. Instead, they wanted to jointly make a measure" (Savickas & Porfelli, 2012, p. 664) considering different cultural setting simultaneously to create *The Career Adapt-Ability Scale* (CAAS). Briefly, the project began with a literature review trying to outline a conceptual framework of career adaptability that distinguished among adaptability, readiness, resources, responses, and results. After that, the team decided to continue in order to construct a measure of career adaptability resources. After, in *viva vocce* meeting the group settled on an international measure, in English, and later translated for other native languages, composed of 44 items. A research methodology group designed the psychometric protocol: a series of hierarchical confirmatory analyses conducted to select 24 items that were relevant across all countries. "The results suggested that the CAAS measures the same construct across all countries, but that the CAAS does not reach scalar invariance, implying that norms have to be developed for each language version" (Rossier, 2015, p.155). This also implies that latent mean level of adaptability cannot be compared across groups.

In order for a better understanding of how and why cultural context affects the construction of life, skilled test developers with personal knowledge of the culture within

which they are operating should develop career instruments. Confronting theory, construct interpretation, and meaning in career management is a global enterprise toward building a fair society. Such confrontation can be translated into methods of research that focus more on integrating ideas on career measurement and on exploring relationships among such measures (Flores & Bike, 2014; Watson, Duarte, & Glavin, 2005). The search for what is common and specific across cultures is a way to achieve contextual meaningfulness.

New perspectives

Van de Vijver and Poortinga (2002) clearly rejected the idea of description of the maximisation of cultural context. They emphasise the importance of abstraction in cross-cultural research together with the minimisation of cultural context. At least "as long as [it does] not make the behaviour studied incomprehensible or irrelevant" (p. 253). They reinforce the idea that only with a deep knowledge of daily contexts (it is assumed that daily context is related with the knowledge of beliefs, values, habits, symbols, expressions) is it possible to reduce culture to a set of centre variables for the construct and proceeding culture-comparative research. This rationale for abstraction is relevant in career psychology field, namely when career assessment is used and evoked. Career assessment was rooted and grew upon the ground of the individual differences and psychometric traditions in psychology, emphasising objective measurement of quantifiable person variables, normative standards, and verifiable realities (Arbona, 2014). Nowadays, based on more integrative theoretical and career assessment approaches (e.g., Busacca & Rehfuss, 2017; Savickas et al., 2009; Stoltz & Barclay, in press), career assessment is not a cumulative process of interpretation of the assessment data; instead, it integrates environment variables that, beyond adding incremental validity to the assessment process, consider also the cultural context. The relationship between a multi approach in counselling and recognising the plurality of contextual/cultural knowledge should take into account (Duarte, 2017). In short, the notion of career assessment

has different characteristics depending upon the cultural specificities. Dropping it out from core variables set (e.g., assessment infused within the career counselling process) diminishes its cultural meaning. In this way, it is possible to compare the effect of psychometric assessment, not reducing assessment (nor evaluation) to abstract generalities but connect it with the way of thinking in the different cultures, and put "the emphasis on contexts and culture" (Guichard & Lenz, 2005, p. 26). The contemporary comprehensive approaches multi-source assessment taking into account the socio-economical context and the cultural of the counselees. This approach implies to go beyond the positivistic cross-cultural psychology and consider new alternative or complementary methodological principles, such as adding using a variety of methods and combining different approaches and sources of information to study psychological group differences.

Final remarks, or the need for integrated approaches

Changing contexts and competitive pressures force the demand for innovation in the field of cross-cultural testing and assessment activities. The importance of links between cultural background and individual's (*idiographic*) assessment seems to be one of the cues. The integration of both etic (in the sense of universal, to a certain extent, universal in their applications, not as an *imposed* etic) (Berry, 1969) and emic approaches or knowledge (providing the utilisation of assessment data) can encourage proactive and innovative forms of testing in an international context (e.g., Fetvadjiev et al., 2015). Therefore, attention needs to be given to the context, both emic-etic conceptualisations of universal variables and culture specific criteria.

The establishment of cultural equivalence can be detected by a clear understanding of methodological procedures: methodological supplies for cross-cultural equivalence have their own rules, in the specification of the notions of equivalence, and the notions of bias. The challenge is the integration of the psychometric or psychological procedures for assessment

in an international context, getting out from a "technocratic" perspective of knowledge (the assumption that tests are universal in their applications) to enter into the development of culture "networks" which encourage exchange and knowledge-sharing.

Critical attention should be given to the equity principle underlying non-discriminatory assessment. The purpose here is to take into account environmental factors within the assessment process in order to avoid any discrimination due to cultural factors. This goal is however difficult to reach knowing those societies themselves do not respect this equity principle. But several techniques, which sometimes imply to adopt a new perspective on testing and assessment, may help career counsellors. Thus, combining a clinical and a psychometric evaluation might help to bring these etic and emic approaches into an assessment process. Moreover, equity in testing and assessment implies of course the use of multicultural counselling competencies by career counsellors (Collins & Arthur, 2010). The available evidence tends to suggest that when using assessment techniques cross-culturally it is essential to develop a thorough understanding of the theoretical literature and of the empirical studies related with the psychological instrument, as well as the understanding of the "new" cultural context, or cultural competence for the understanding of the individual's needs.

Incorporating Vygotsky's (1987) historical conception of dynamic testing into modern psychology, in contrast with conventional or psychometric approaches used only to provide diagnostic information, an alternative and more integrative approach could be considered. Regarding the consequences of globalisation, in what seems to be the new societal needs, it is mandatory to recognise the desirability of a greater proximity to culture in testing, that is to say, the use of testing and assessment considering different cultural background. Cross-national evidence is an added value to test validation and could be used in the understanding of multi-faceted profile more adapted to local, regional, or countries situations.

The Work Importance Study project opened a gate for testing in an international context and demonstrated the applicability of measures developed across cultures. 21st century, and following the path of Donald Super, an international team of vocational psychologists, crafted an operational definition for the linguistic conception of career resources. The future? Keep the gate open, considering scientific pertinence, ecological meaningfulness, societal needs, comparing results obtained with cultural diversified populations, developing non-discriminatory assessment devices. Testing and assessment in an international context focused on the appropriateness of the measures chosen on the basis of psychometric and cultural criteria; testing and assessment linked to intervention and encompassing the integration between the individual and the ecological context in order to help with knowledgeable counselling; testing and assessment as a form of testing hypotheses with cultural representative samples. In sum, testing and assessment in an international context is viewed as a teamwork task to achieve success in what concerns the analytical (analysing, comparing, and evaluating results), the practical (applying, utilising), and the creative (inventing and designing comprehensive research) issues in the field of career guidance.

Acknowledgment

References

American Psychological Association (1990). *Guidelines for providers of psychological services to ethnic, linguistic and culturally diverse populations*. Washington, DC: Author, Board of Ethnic Minority Affairs, Task Force on the Delivery of Service to Ethnic Minority Populations.

Arbona, C. (2014). Career assessment. In F. T. L. Leong, L. Comas-Diaz, G. C. Nagayama Hall, V. C. McLoyd, & J. E. Trimble (Eds.), *APA handbook of multicultural psychology, Vol. 2: Applications and training* (pp. 215-230). Washington, DC: American Psychological Association. doi:10.1037/14187-013

Armstrong, P. I., Hubert, L., & Rounds, J. (2003). Circular unidimensional scaling: A new look at group differences in interest structure. *Journal of Counseling Psychology, 50,* 297-308. doi:10.1037/0022-0167.50.3.297

Atitsogbe, K. A., Moumoula, I. A., Rochat, S., & Rossier, J. (submitted). Vocational interests and career indecision in Switzerland and Burkina Faso using the personal globe nventory and the career decision-making difficultires questionnaire. *Journal of Vocational Behavior*.

Association for Assessment in Counseling (2003). *Standards for multicultural assessment* (2nd ed.). Retrieved September 20, 2006, from http://aac.ncat.edu/resources/documents

Baruth, L. G., & Manning, M. L. (1992). *Multicultural education and children and adolescents*. Needham Heights, MA: Allyn & Bacon.

Berry, J. (1969). On cross-cultural comparability. *International Journal of Psychology, 4,* 119-128. doi:10.1080/00207596908247261

Blustein, D. (2006). *The psychology of working*. Mahwah, NJ: Lawrence Erlbaum Associates.

Busacca, L. A., & Rehfuss, M. C. (Eds.). (2017). *Postmodern career counseling: A handbook of culture, context, and cases*. Alexandria, VA: American Counseling Association.

Butcher, J.N. (1996). Translation and adaptation of the MMPI-2 for international use. In J. N. Butcher (Ed.), *International adaptations of the MMPI-2* (pp.26-43). Minneapolis, MN: University of Minnesota Press.

Byrne, B. M. (2016). Multigroup comparisons: Testing for measurement, strutural, and latent mean equivalence. In F. T. L. Leong, D. Bartram, F. M. Cheung, K. F. Geisinger, & D. Iliescu (Eds.), *The ITC international handbook of testing and assessment* (pp. 377-394). Oxford, UK: Oxford University Press.

Cattell, R. (1940). A culture-free intelligence test, I. *Journal of Educational Psychology, 31*, 176-199. doi:10.1037/h0059043

Cattell, R., & Cattell, A. (1963). *Culture fair intelligence test*. Champaing, IL: Institute for Personality and Ability Testing.

Collins, S., & Arthur, N. (2010). Culture-infused counselling: A model for developing multicultural competence. *Counselling Psychology Quarterly*, 23, 217-233. doi:10.1080/09515071003798212

Cronbach, L. (1990). *Essentials of psychological testing* (5th ed.). New York, NY: Harper Collins.

du Toit, R., & De Bruin, G. P. (2002). The structural validity of Holland's model of vocational personality types for young Black South African men and women. *Journal of Career Assessment*, 10, 62-77. doi:10.1177/1069072702010001004

Duarte, M. E. (1993). *Preocupações de carreira, valores e saliência das actividades em adultos empregados*. [Career concerns, values and role salience in employed adults]. Doctoral dissertation, University of Lisbon, Portugal.

Duarte, M. E. (1995). Career concerns, values and role salience in employed men. *The Career Development Quarterly, 43*, 338-349. doi:10.1002/j.2161-0045.1995.tb00438.x

Duarte, M. E. (2005). Assessment and "The global Research Village": Adapting

Psychological instruments. *International Journal for Educational and Vocational

Guidance, 5*, 163-173. doi:10.1007/s10775-005-8797-y

Duarte, M. E. (2017). Career counseling research-practice disparities: what we know and

what we need to know. *South African Journal of Education, 37*, 1-7.

doi:10.15700/saje.v37n4a1486

Duarte, M. E., & Rossier, J. (2008). Testing and assessment in an international context :

Cross- and multi-cultural issues. In J. Athanasou & R. Van Esbroeck (Eds.),

International handbook of career guidance (pp. 489-510). New York, NY: Springer

Science. doi:10.1007/978-1-4020-6230-8_24

Einarsdóttir, S., Eyjólfsdóttir, K. Ó., & Rounds, J. (2013). Development of indigenous basic

interest scales: Re-structuring the Icelandic interest space. *Journal of Vocational

Behavior, 82*, 105-115. doi:10.1016/j.jvb.2013.01.001

Fetvadjiev, V. H., Meiring, D., van de Vijver, F. J. R., Nel, J. A., & Hill, C. (2015). The

South African personality inventory (SAPI): A culture-informed instrument for the

country's main ethnocultural groups. *Psychological Assessment, 27*, 827-837.

doi:10.1037/pas0000078

Flores, L. Y., & Bike, D. H. (2014). Multicultural career counseling. In F. T. L. Leong, L.

Comas-Diaz, G. C. Nagayama Hall, V. C. McLoyd, & J. E. Trimble (Eds.), *APA

handbook of multicultural psychology, Vol. 2: Applications and training* (pp. 403-417).

Washington, DC: American Psychological Association. doi:10.1037/14187-023

Fouad, N. A., & Byars-Winston, A. M. (2005). Cultural context of career choice: Meta-

analysis of race/ethnicity differences. *The Career Development Quarterly, 53*, 223-233.

doi:10.1002/j.2161-0045.2005.tb00992.x

Gjersing, L., Caplehorn, J. R. M., & Clausen, T. (2010). Cross-cultural adaptation of research instruments: Language, setting, time and statistical considerations. *BMC Medical Research Methodology, 10*, 13. doi:10.1186/1471-2288-10-13

Goldman, L. (1992). Qualitative assessment: An approach for counselors. *Journal of Counseling & Development, 70*, 616-621. doi:10.1002/j.1556-6676.1992.tb01671.x

Gopaul-McNicol, S., & Armour-Thomas, E. (2002). *Assessment and culture: Psychological tests with minority populations*. San Diego, CA: Academic Press.

Guichard, J., & Lenz, J. (2005). Career theory from an international perspective. *The Career Development Quarterly, 54*, 17-28. doi:10.1002/j.2161-0045.2005.tb00137.x

Hambleton, R. K. (1993). Translating achievement tests for use in cross-national studies. *European Journal of Psychological Assessment, 9*(1), 57-68.

Hambleton, R. K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment, 17*, 164-172. doi:10.1027//1015-5759.17.3.164

Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. Hambleton, P. Merenda & C. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum Associates.

Hambleton, R. K., Merenda, P., & Spielberger, C. (2005, Eds.). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology, 1*(2), 1-30. Retrieved from http://www.jattjournal.com/index.php/atp/article/view/48345/39215

Hartung, P. (2005). Integrated career assessment and counseling. Mindsets, models, and methods. In W. B. Walsh, & M. Savickas (Eds.), *Handbook of Vocational Psychology* (3rd ed., pp. 371-395). Mahwah, NJ: Lawrence Erlbaum Associates.

Hedrih, V., Stošić, M., Simić, I., & Ilieva, S. (2016). Evaluation of the hexagonal and spherical model of vocational interests in the young people in Serbia and Bulgaria. *Psihologija*, *49*, 199-210. doi:10.2298/PSI1602199H

Heim, E., Scholten, S., Maercker, A., Xiu, D., Cai, D., Gao, Z. H., … Margraf, J. (2017). Students' value orientations in contemporary China: Analysis of measurement invariance and latent mean differences in comparison with students from Germany and Russia. *Journal of Cross-Cultural Psychology*, *48*, 511-531. doi: 10.1177/0022022117696800

International Test Commission (2010). *International Test Commission guidelines for translating and adapting tests*. Retrived from http://www.intestcom.org

Jensen, A. (1980). *Bias in mental testing*. New York, NY: Free Press.

Johnson, T. P., Shavitt, S., & Holbrook, A. L. (2011). Survey response style across cultures. In D. Matsumoto & F. J. R. van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 130-175). Cambridge, UK: Cambridge University Press.

Kitayama, S. (2002). Culture and Basic Psychological Processes—Toward a System View of Culture: Comment on Oyserman et al. (2002). *Psychological Bulletin, 128*, 89–96. doi:10.1037//0033-2909.128.1.89

Laher, S., & Cockcroft, K. (2017). Moving from culturally biased to culturally responsive assessment practices in low-resource, multicultural settings. *Professional Psychology: Research and Practice*, *48*, 115-121. doi:10.1037/pro0000102

Leong, F. T. T., Bartram, D., Cheung, F., Geisinger, K. F., & Iliescu, D. (Eds.). (2016). *The ITC international handbook of testing and assessment*. Oxford, UK: Oxford University Press.

Leong, F. T. L., & Brown, M. T. (1995). Theoretical issues in cross cultural career development: Cultural validity and cultural specificity. In W. B. Walsh & S. H. Osipow (Eds.), *Handbook of vocational psychology* (2nd ed., pp. 143-180). Hillsdale, NJ: Erlbaum.

Leong, F. T. L., & Pearce, M. (2014). Indigenous models of career development and vocational psychology. In G. Arulmani, A. J. Bakshi, F. T. L. Leong, & A. G. Watts (Eds.), *Handbook of career development: International perspectives* (pp. 67-79). New York, NY: Springer. doi:10.1007/978-1-4614-9460-7_5

Leong, F. T. L., & Walsh, W. B. (2012). Guest editors' introduction to the special issue. *Journal of Vocational Behavior, 80*, 659-660. doi:10.1016/j.jvb.2012.03.003

Long, L., & Tracey, T. J. (2006). Structure of RIASEC scores in China: A structural meta-analysis. *Journal of Vocational Behavior, 68*, 39-51. doi:10.1016/j.jvb.2005.01.002

Long, L., Watanabe, N., & Tracey, T. J. (2006). Structure of interests in Japan: Application of the personal globe inventory occupational scales. *Measurement and Evaluation in Counseling and Development, 38*, 222-235. doi:10.1080/07481756.2006.11909784

Marsella, A. J., & Leong, F. T. L. (1995). Cross-cultural issues in personality and career assessment. *Journal of Career Assessment, 3*, 202-218. doi:10.1177/106907279500300207

Matsumoto, D., & Van de Vijver, F. J. R. (Eds.). (2011). *Cross-cultural research methods in psychology*. Cambridge, UK: Cambridge University Press.

McCrae, R. R. (2017). The five-factor model across cultures. In A. T. Church (Eds.), *The Praeger handbook of personality across cultures* (Vol. 1, pp. 47-71), Santa Barbara, CA: Praeger Publisher.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525-543. doi:10.1007/BF02294825

Metz, A. J., Fouad, N., & Ihle-Helledy, K. (2009). Career aspirations and expectations of college students: Demographic and labor market comparisons. *Journal of Career Assessment*, *17*, 155-171. doi:10.1177/1069072708328862

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Routledge.

Munroe, R. H., Munroe, R. L., & Whiting, B. B. (Eds.). (1981). *Handbook of cross-cultural human development*. New York, NY: Garland.

Naglieri, J. A. (2016). *Naglieri nonverbal ability test* (3rd ed.). San Antonio, TX: Pearson.

Oakland, T. (2004). Use of educational and psychological tests internationally. *Applied Psychology: An International Review, 53*, 157-172. doi:10.1111/j.1464-0597.2004.00166.x

OECD (2017). *International migration outlook 2017*. Paris, France: OECD Publishing. doi:10.1787/migr_outlook-2017-en

Osborn, D. S. (2012). An international discussion about cross-cultural career assessment. *International Journal for Educational and Vocational Guidance*, *12*, 5-16. doi:10.1007/s10775-012-9220-0

Raykov, T., Marcoulides, G. A., & Li, C.-H. (2012). Measurement invariance for latent constructs in multiple populations: A critical view and refocus. *Educational and Psychological Measurement*, *72*, 954-974. doi:10.1177/0013164412441607

Rossier, J. (2005). A review of cross-cultural equivalence of frequently used personality inventories. *International Journal for Educational and Vocational Guidance, 5*, 175-188. doi:10.1007/s10775-005-8798-x

Rossier, J. (2015). Career adaptability and life designing. In L. Nota & J. Rossier (Eds.), *Handbook of life design* (pp.153-167). Boston, MA: Hogrefe.

Rossier, J., Aluja, A., Blanch, A., Barry, O., Hansenne, M., Carvalho, A., … Karagonlar, G., (2016). Cross-cultural generalizability of the alternative five-factor model using the Zuckerman-Kuhlman-Aluja personality questionnaire. *European Journal of Personality*, *30*, 139-157. doi:10.1002/per.2045

Rossier, J., & Fiori, M. (in press). Career assessment in Europe: Overview and current trends. In K. B. Stoltz & S. R. Barclay (Eds.), *A counselor's guide to comprehensive career assessments* (7th ed.). Broken Arrow, OK: NCDA.

Rossier, J., Ouedraogo, A., & Dahourou, D. (2017). Personality structure and assessment in French-speaking African Cultures. In A. T. Church (Eds.), *The Praeger handbook of personality across cultures* (Vol. 1, pp. 73-103), Santa Barbara, CA: Praeger Publisher.

Rossier, J., Ouedraogo, A., Dahourou, D., Verardi, S., & Meyer de Stadelhofen, F. (2013). Personality and personality disorders in urban and rural Africa: Results from a field trial in Burkina Faso. *Frontiers in Psychology*, *4*, 79. doi:10.3389/fpsyg.2013.00079

Rounds, J., & Tracey, T. J. (1996). Cross-cultural structural equivalence of RIASEC models and measures. *Journal of Counseling Psychology, 43*, 310-329. doi:10.1037/0022-0167.43.3.310

Ryan, J. M., Tracey, T. J. G., & Rounds, J. (1996). Generalizability of Holland's structure of vocational interest across ethnicity, gender, and socioeconomic status. *Journal of Counseling Psychology, 43*, 330-337. doi:10.1037/0022-0167.43.3.330

Savickas, M. (1989). Career-style assessment and counseling. In T. Sweeney (Ed.), *Adlerian counseling: A practical approach for a new decade* (3rd ed., pp. 289-320). Muncie, IN: Accelerated Development Press.

Savickas, M. (2004). The theory and practice of career construction. In S. Brown & R. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (pp. 42-70). Hoboken, NJ: Wiley.

Savickas, M. L., Nota, L., Rossier, J., Dauwalder, J.-P., Duarte, E., Guichard, J., … van Vianen, A. E. M. (2009). Life designing: A paradigm for career construction in the 21st century. *Journal of Vocational Behavior, 75*, 239-250. doi:10.1016/j.jvb.2009.04.004

Savickas, M. L., & Porfeli, E. (2012). Career adapt-abilities scale: Construction, reliability, and measurement equivalence across 13 countries. *Journal of Vocational Behavior, 80*, 661-673. doi:10.1016/j.jvb.2012.01.001

Savickas, M., Van Esbroeck, R., & Herr, E. (2005). The internationalization of educational and vocational guidance. *The Career Development Quarterly, 54*, 77-85. doi:10.1002/j.2161-0045.2005.tb00143.x

Schwabe, F., von Davier, A. A., & Chalhoub-Deville, M. (2016). Language and culture in testing. In F. T. T. Leong, D. Bartram, F. Cheung, K. F. Geisinger, & D. Iliescu (Eds.), *The ITC international handbook of testing and assessment* (pp. 300-317). Oxford, UK: Oxford University Press.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1-65). San Diego, CA: Academic Press.

Schwartz, S. H. (2017). Individual values across cultures. In A. T. Church (Eds.), *The Praeger handbook of personality across cultures* (Vol. 2, pp. 121-152), Santa Barbara, CA: Praeger Publisher.

Schwartz, S. H., & Bilsky, W. (1990). Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *Journal of Personality and Social Psychology*, *58*, 878-891. doi:10.1037/0022-3514.58.5.878

Schwartz, S. H., & Sagiv, L. (1995). Identifying culture-specifics in the content and structure of values. *Journal of Cross-Cultural Psychology*, *26*, 92-116. doi:10.1177/0022022195261007

Stoltz, K. B., & Barclay, S. R. (Eds.). (in press). *A counselor's guide to comprehensive career assessments* (7th ed.). Broken Arrow, OK: NCDA.

Strauss, M. E., & Smith, G. T. (2009). Construct Validity: Advances in Theory and Methodology. *Annual Review of Clinical Psychology*, *5*, 1–25. http://doi.org/10.1146/annurev.clinpsy.032408.153639

Super, D. (1990). A life-span, life-space approach to career development. In D. Brown & L. Brooks (Eds.), *Career choice and development* (2nd ed., pp. 197-261). San Francisco, CA: Jossey-Bass.

Super, D., & Šverko, B. (1995). *Life roles, values and careers*. San Francisco, CA: Jossey-Bass.

Super, D., Thompson, A., & Lindeman. R. (1985). *The Adult Career Concerns Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Tracey, T. J. G. (2002). Personal globe inventory: Measurement of the spherical model of interests and competence beliefs. *Journal of Vocational Behavior*, *60*, 113-172. doi:10.1006/jvbe.2001.1817

Van de Vijver, F. J. R. (2000). The nature of bias. In R. H. Dana (Ed.), *Handbook of cross-cultural and multicultural personality assessment* (pp. 87-106). Mahwah, NJ: Lawrence Erlbaum Associates.

Van de Vijver, F. J. R. (2015). Methodological aspects of cross-cultural research. In M. Gelfand, C-y Chiu, & Y-y Hong (Eds.), *Handbook of advances in culture and psychology* (vol.5, pp. 101-160). New York, NY: Oxford University Press.

Van de Vijver, F. J. R., & Hambleton, R. (1996). Translating tests: some practical guidelines. *European Psychologist, 1*, 89-99. doi:10.1027/1016-9040.1.2.89

Van de Vijver, F. J. R., & Leung, K. (2011). 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. 17-45). Cambridge, UK: Cambridge University Press.

Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage.

Van de Vijver, F. J. R., & Matsumoto, D. (2011). Introduction to the methodological issues associated with cross-cultural research. In D. Matsumoto & F. J. R. van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 1-14). Cambridge, UK: Cambridge University Press.

Van de Vijver, F. J. R., & Poortinga, Y. H. (2002). On the study of culture in developmental science. *Human Development, 45*, 246-256. doi:10.1159/000064985

Vygotsky, L. (1987). *The collected works of L. S. Vygotsky, Vol. 1*. New York, NY: Plenum.

Warner, O., & Campbell, D. T. (1970). Translating, working through interpreters, and the problem of decentering. In R. Naroll & R. Cohen (Eds.), *A Handbook of cultural anthropology* (pp. 398-419). New York, NY: American Museum of Natural History.

Watson, M., Duarte, M. E., & Glavin, K. (2005). Cross-cultural perspectives on career assessment. *The Career Development Quarterly, 54*, 29-35. doi:10.1002/j.2161-0045.2005.tb00138.x

Wechsler, D., & Naglieri, J. A. (2006). *Wechsler nonverbal scale of ability (WNV)*. San

Antonio, TX: Pearson.

Wonderlic Inc. (1983). *Wonderlic Personnel Test Manual*. Northfield, IL: Author.

Yasui, M. (2015. The Cultural Ecogram: A Tool for Enhancing Culturally Anchored Shared

Understanding in the Treatment of Ethnic Minority Families. *Journal of Ethnic &

Cultural Diversity in Social Work*, *24*, 89-108. doi:10.1080/15313204.2014.991980

Zercher, F., Schmidt, P., Cieciuch, J., & Davidov, E. (2015). The comparability of the

universalism value over time and across countries in the European Social Survey:

Exact vs. approximate measurement invariance. *Frontiers in Psychology*, *6*, 733.

doi:10.3389/fpsyg.2015.00733