

Cross-Cultural Adaptation and Validation of Psychological Instruments: Some Considerations¹

Juliane Callegaro Borsa²
Bruno Figueiredo Damásio
Denise Ruschel Bandeira

Universidade Federal do Rio Grande do Sul, Porto Alegre-RS, Brazil

Abstract: The adaptation of psychological instruments is a complex process that requires a high methodological rigor. Because there is no consensus in the literature about its steps, this article discusses some essential aspects regarding the cross-cultural adaptation of psychological instruments and proposes guidelines to the researchers about the different steps of this process. Some considerations regarding the validation of the adapted instrument are also presented. In this stage, we discuss some aspects regarding the factorial structure of the instrument, which might be evaluated through statistical procedures, such as exploratory and confirmatory factor analysis. More than that, the authors provide some guidelines to the validation of psychological instruments in different cultures.

Keywords: translating, adaptation, psychological testing, psychometrics

Adaptação e Validação de Instrumentos Psicológicos entre Culturas: Algumas Considerações

Resumo: A adaptação de instrumentos psicológicos é um processo complexo que requer elevado rigor metodológico. Por não haver consenso na literatura sobre suas etapas, o presente artigo discute alguns aspectos essenciais concernentes à adaptação transcultural de instrumentos psicológicos e propõe diretrizes aos pesquisadores sobre os diferentes passos desse processo. São apresentadas, também, algumas considerações referentes à validação do instrumento adaptado. Nesta etapa, são discutidos os aspectos referentes à estrutura fatorial do instrumento, a qual requer avaliação por meio de procedimentos estatísticos, como análises fatoriais exploratórias e confirmatórias, sendo fornecidas algumas diretrizes gerais para a validação de instrumentos psicológicos em diferentes culturas.

Palavras-chave: tradução, adaptação, testes psicológicos, psicometria

Adaptación y Validación de Instrumentos Psicológicos entre Culturas: Algunas Consideraciones

Resumen: La adaptación de instrumentos psicológicos es un proceso complejo que requiere bastante rigor metodológico. Ya que no hay consenso sobre sus etapas, el presente artículo discute algunos aspectos esenciales sobre la adaptación transcultural de instrumentos psicológicos y propone directrices a los investigadores sobre los diferentes pasos de este proceso. Son presentadas, también, algunas consideraciones referentes a la validación del instrumento adaptado. En esta etapa, son discutidos aspectos referentes a la estructura factorial del instrumento, la cual debe ser evaluada mediante procedimientos estadísticos como el análisis factorial exploratorio y confirmatorio. Además, se incluyen algunas directrices para la validación de instrumentos psicológicos en culturas diversas.

Palabras clave: traducción, adaptación, testes psicológicos, psicometría

The adaptation of psychological instruments is a complex task that requires careful planning regarding its content maintenance, psychometric properties, and general validity for the intended population (Cassepp-Borges, Balbinotti, & Teodoro, 2010). During this process, one must provide both the evidences of the semantic equivalence of the items and the adequate psychometric properties of the

new version of the instrument (International Test Commission [ITC], 2010). It is also important that the adaptation of an instrument possesses a cultural fit, that is, its preparation for use in different cultural contexts (Beaton, Bombardier, Guillemin, & Ferraz, 2000; Hambleton, 2005; Sireci, Yang, Harter, & Ehrlich, 2006).

Since 1992, the International Test Commission (ITC) has been working to propose guidelines for the cross-cultural translation and adaptation of psychological instruments (ITC, 2010). The terms "adaptation" and "translation" are distinct, and the former has been used most often because it includes all the processes concerning the cultural fit of the instrument beyond mere translation (Hambleton, 2005).

¹ Support: National Council for Scientific and Technological Development (CNPq).

² Correspondence address:

Juliane Callegaro Borsa. Universidade Federal do Rio Grande do Sul. Rua Ramiro Barcelos, 2600, sala 120. CEP 90.035-003. Porto Alegre-RS, Brazil. E-mail: juliborsa@gmail.com

Translation is merely the first stage of the adaptation process. When adapting an instrument, cultural, idiomatic, linguistic and contextual aspects concerning its translation should be considered (Hambleton, 2005). Once the instrument is adapted, studies between different populations that compare the characteristics of individuals in different cultural contexts may be conducted. Accordingly, research on the adaptation of instruments has placed great emphasis on comparing results through studies that use different samples (Gjersing, Caplehorn, & Clausen, 2010; Hambleton, 2005).

The process of adapting an existing instrument, rather than developing a new one that is specifically for the target population, has considerable advantages. By adapting an instrument, the researcher is able to compare data from different samples and from different backgrounds, which enables greater fairness in the evaluation because the same instrument assesses the construct based on the same theoretical and methodological perspectives. The use of adapted instruments naturally enables a greater ability to generalize and also enables one to investigate differences within an increasingly diverse population (Hambleton, 2005; Vivas, 1999).

The present article reviews a few key aspects concerning the cross-cultural adaptation of psychological instruments and proposes guidelines on the different stages of this process to researchers. The topics will be presented according to the authors' proposal for conducting the adaptation process. In general, the literature indicates that instrument adaptation entails five essential stages: (1) instrument translation from the source language into the target language, (2) synthesis of the translated versions, (3) analysis of the synthesized version by expert judges, (4) back translation, and (5) a pilot study (Hambleton, 2005; Sireci et al., 2006). However, we understand that there are key aspects that are important in the process of adapting the new version of an instrument that are not included in these stages, such as the conceptual evaluation of items by the target population, and a discussion with the original instrument's author regarding the proposed changes in the new version of the instrument.

Therefore, we present our proposal for the adaptation of instruments based on six stages: (1) instrument translation from the source language into the target language, (2) synthesis of the translated version, (3) a synthesis evaluation by expert judges, (4) instrument evaluation by the target population, (5) back translation, and (6) a pilot study. In addition, we will discuss a seventh stage that is normally not included in the adaptation process but that we deem important to confirm whether the instrument structure is stable when compared to the original instrument. This stage involves the evaluation of the factorial structure of the instrument, which is accomplished by statistical procedures, including exploratory and confirmatory factor analyses. We further discuss the procedures regarding the validation of

instruments for cross-cultural studies, in which the instrument is tested in different cultures to verify the stability of its structure and parameters when applied to different cultural groups and contexts.

Stages of the Translation Process and the Adaptation of Instruments

Instrument Translation into the New Language

When adapting an instrument, one must first consider its translation from the source language into the target language, that is, the language with which the new version will be used. This is a complex process and requires tremendous care to ensure that the final version is not only suitable for the new context but is also consistent with the original version.

Previous literature emphasizes the need to avoid the literal translation of items (Hambleton, 1994, 2005) because that often results in incomprehensible statements or rather limited target language fluency. Therefore, an appropriate translation requires a balanced treatment of linguistic, cultural, contextual, and scientific information (Tanzer, 2005).

The research consensus in this area suggests that independent, bilingual translators should be summoned to adapt the items into the new language (Beaton et al., 2000; Gudmundsson, 2009; Hambleton, 2005; ITC, 2010). Although a single translator was previously believed to be suitable for the completion of the translation process, the presence of at least two bilingual translators is now recommended for completing this process, thereby minimizing the risk of linguistic, psychological, cultural, and both theoretical and practical understanding biases (Cassepp-Borges et al., 2010).

Many suggestions for translation focus on the quality of the translators. For example, Hambleton (1994, 2005) argues that translators should be fully proficient in both languages of interest and familiar with the cultures associated with the respective languages of each group. Beaton et al. (2000) advocate that translators must be fluent in the source language of the instrument and native in the target language. Such characteristics enable the translation process to consider the nuances of the language for which the instrument is intended, which ensures a greater cultural fit of the adaptation process.

For other authors, translators are expected to understand the construct that is being assessed and to have both scientific writing skills and familiarity with the subject (Cassepp-Borges et al., 2010; Hambleton, 1994, 2005; ITC, 2010). For Beaton et al. (2000), one of the translators should be familiar with the assessed construct, while a second translator should not be aware of the translation goals. The first translator's adaptation tends to provide a higher

scientific similarity of the instrument, which delivers a higher equivalence from a psychometric perspective. Conversely, the second translator's adaptation tends to show a lower probability of deviation concerning the meaning of items. The second translator tends to offer a version that best reflects the language used by the target population because he or she is less influenced by the academic purpose of the translation.

Synthesis of the Translated Versions

After the process of instrument translation from the source language into the target language, the researcher should have at least two versions of the translated instrument. At this stage, the process of summarizing both versions begins. Summarizing the versions of an instrument means that the researcher compares the different translations and assesses their semantic, idiomatic, conceptual, linguistic and contextual differences, with the sole purpose of creating a single version. During this process, it is common to identify two possible sources of complications: (1) complex translations that may hinder the understanding of the population for whom the instrument is intended or (2) overly simplistic translations that underestimate the item content. Inappropriate choices are identified and resolved through discussions among the judges (experts in the area that the instrument proposes to assess) and researchers responsible for adapting the instrument.

The evaluation of different translations of an instrument should be conducted for each item separately. Throughout this process, the committee (judges and authors) should assess the compatibility between the translated versions and the original instrument in the following four different areas: (1) semantic equivalence, which aims to assess whether the words have the same meaning, whether the item has more than one meaning, and whether there are grammatical errors in the translation; (2) idiomatic equivalence, which refers to assessing whether the items from the original instrument that are difficult to translate were changed into an equivalent expression that has not changed the cultural meaning of the item; (3) experiential equivalence, which refers to noting whether a particular item is applicable in the new culture and, if not, replacing it with an equivalent item; and (4) conceptual equivalence, which seeks to assess whether a given term or expression, even if properly translated, assesses the same aspect in different cultures. If the translated versions are flawed in one or more of these aspects, the committee may propose a new translation that is better suited for the instrument characteristics and the cultural context in which it will be used. In such cases, the participation of researchers is crucial because they should have sufficient knowledge about the topic assessed by the instrument; therefore, the researchers may resolve theoretical doubts about the items and help decide on the best expressions to use.

The choice of which version to use must be made through consensus among the judges, and never by imposition (Gjersing et al., 2010). When possible, an external observer should be prompted to transcribe the entire synthesis process, especially regarding the choice of items to be used (Beaton et al. 2000). This transcription provides a qualitative overview of the process to the researcher. At the end of this stage, the researcher will hold a single version of the instrument, which may include items translated by one or more than one translators (Gudmundsson, 2009).

Evaluation of the Synthesized Version by Experts

After the synthesis of translated versions has been finished, the researcher should still rely on the help of a committee of either experts in the area of psychological evaluation or on those with specific knowledge of what the instrument assesses. These experts will assess other important aspects, such as the structure, layout, instrument instructions, and both the scope and adequacy of expressions contained in the items. The experts will then consider, for example, whether the terms or expressions may be generalized for different contexts and populations (that is, different regions of a given country) and whether the expressions are a good fit for the population for whom the instrument is intended. Aspects of the instrument layout will also be analyzed because they are as indispensable as the linguistic aspects of the items, especially regarding the instruments to be used on specific populations, including children and the elderly. The clarity of the content, the suitability of font formats and sizes, the arrangement of information on the instrument, *inter alia*, are also analyzed.

One example concerns the adaptation study of a spiritual-religious coping scale, which was conducted by Panzini and Bandeira (2005). This instrument assesses how individuals use faith to cope with stress. During the adaptation process, researchers submitted the instrument to a group of experts, or religious leaders, on the topic of "spiritual-religious coping." It was critical to interview leaders of different religious institutions because the number of religions in Brazil was larger than in the instrument's country of origin. One of the key contributions made by these judges was to determine to what extent the proposed terms were appropriate and could be generalized for many different religions.

Translation, synthesis and evaluation of the translated version are the first steps in the instrument's adaptation into a new culture. After completing these stages, the first version of the instrument will be ready for the next stage: instrument evaluation by the target population.

Evaluation by the Target Population

This stage of the process aims to verify whether the items, the response scale and the instructions are

comprehensible for the target population. Thus, this procedure aims to investigate whether the instructions are clear, whether the terms found in the items are appropriate, whether the expressions correspond to those used by the group, and other aspects. The subjects who participate in this step may vary depending on the characteristics of the respondents for whom the instrument is intended. For example, when considering a self-report instrument designed to assess the aggressive behavior of children, the instrument must be presented to a group of children so that researchers can confirm the items' clarity and the extent to which the expressions are representative of the vocabulary commonly used by the group. Accordingly, we suggest that children of different ages (within the intended age group) evaluate this instrument, in addition to residents of different regions (once validated, the instrument can be applied to different populations in different areas of the country).

We must stress that during the evaluation of the target population any statistical procedure is conducted, but rather it is evaluated solely the item appropriateness and instrument structure as a whole (whether the terms are clear, appropriate, and well-written). When a given item is not clear, for example, the respondent is encouraged to provide synonyms that best exemplify the vocabulary of the group for whom the instrument is intended. At this stage, the respondent may be prompted to read the questions aloud and to give a brief explanation of the meaning of each item. The instrument may also be administered so that respondents fill it out and then start a discussion about the clarity of each item, suggesting changes, if necessary. The stage of evaluation by the target population may be conducted one or more times, depending on the need and the complexity of the instrument to be adapted.

Back-Translation

Back-translation is also suggested as an additional quality control check (Sireci et al., 2006). From our perspective, this procedure must follow all semantic and idiomatic adjustment procedures because the instrument must be "ready" for final evaluation by the original author. Back-translation refers to translating the synthesized and revised versions of the instrument into the source language. Its aim is to evaluate the extent to which the translated version reflects the item content of the original version.

According to Beaton et al. (2000), back-translation must be performed by at least two translators other than those who performed the first translation. Several authors have been cautious about the use of back-translation (Gudmundsson, 2009; Hambleton, 1993; Van de Vijver & Leung, 1997). For these authors, the process of back-translation may focus too heavily on grammatical aspects at the expense of contextual aspects. Furthermore, they claim that back-translation disregards what has been advocated thus far: that by adapting an instrument, cultural,

idiomatic, linguistic, or contextual aspects need to be considered. The purpose of back-translation, however, should not be a literal interpretation of the original version and the translated versions. Instead, the back-translation process should be used as a tool to identify words that were not clear in the target language and to identify inconsistencies or conceptual errors in the final version (Beaton et al., 2000). Back-translation may also be used as a practical tool so that the researcher who is adapting the instrument may communicate with the author of the original instrument. When the author has access to the back-translated version of the instrument, the author may state whether the items share the same meanings as those of the original items.

An example that demonstrates the usefulness of communication between authors following the back-translation process is the Inventory of Personality Organization adaptation procedure, which was conducted by Oliveira and Bandeira (2011). During this process, there was a disagreement about a certain item, which was (in the original English) "I am a hero worshipper even if I am later found wrong in my judgment," specifically because the equivalent expression for the term "hero worshipper" ("adorador de heróis") has no clear meaning in Brazilian Portuguese. The item was translated into Portuguese as "Eu idolatro algumas pessoas, mesmo que depois eu me dê conta de que estava enganado" ("I idolize some people, even if I subsequently realize that I was wrong") and back-translated into English as "I idolize some people, even after realizing that I was wrong about them." The original author, however, disagreed on this back-translation. Because of this complication, there were several email exchanges between the authors to discuss the actual meaning of the original expression "hero worshipper." The author of the original instrument understood that the term referred to a mechanism of idealization and was then told that the literal translation into Portuguese would not be appropriate and that the term "idolstrar" ("to idolize") would have the same connotation of "idealização" ("idealization"). It was further argued that the expression "Eu idolatro algumas pessoas" ("I idolize some people") was inspired by the Argentinean version of the instrument, which had already been approved by the same author.

As shown by the previous example, back-translation does not imply that an item must remain literally identical to the original but rather it must maintain a conceptual equivalence. Therefore, the authors must be aware of the possibility of such approximations and consider the meaning of the item in its appropriate cultural context.

Pilot Study

Before claiming that a new instrument is ready for application, one must perform a pilot study. The pilot study refers to a previous application of the instrument in a small sample

that reflects the sample/target-population characteristics (Gudmundsson, 2009). Once again, the appropriateness of items regarding their meaning and difficulty, in addition to instructions for conducting the test, should be assessed during this process. After considering the modifications suggested in the first pilot study, a second pilot study (or as many as needed) is necessary to assess whether the instrument is ready to be used.

To avoid any type of bias, the changes suggested by the pilot study (or studies) should be implemented with the help

of a committee of experts and should never be performed solely by the field researcher. As can be observed, the adaptation process of an instrument into a new culture consists of different stages, which, as suggested by different authors (Beaton et al., 2000; Gjersing et al., 2010; Hambleton, 2005), are essential for conducting the process adequately.

Figure 1 shows a methodology outline proposed by the authors for translating and adapting psychological instruments into different cultures.

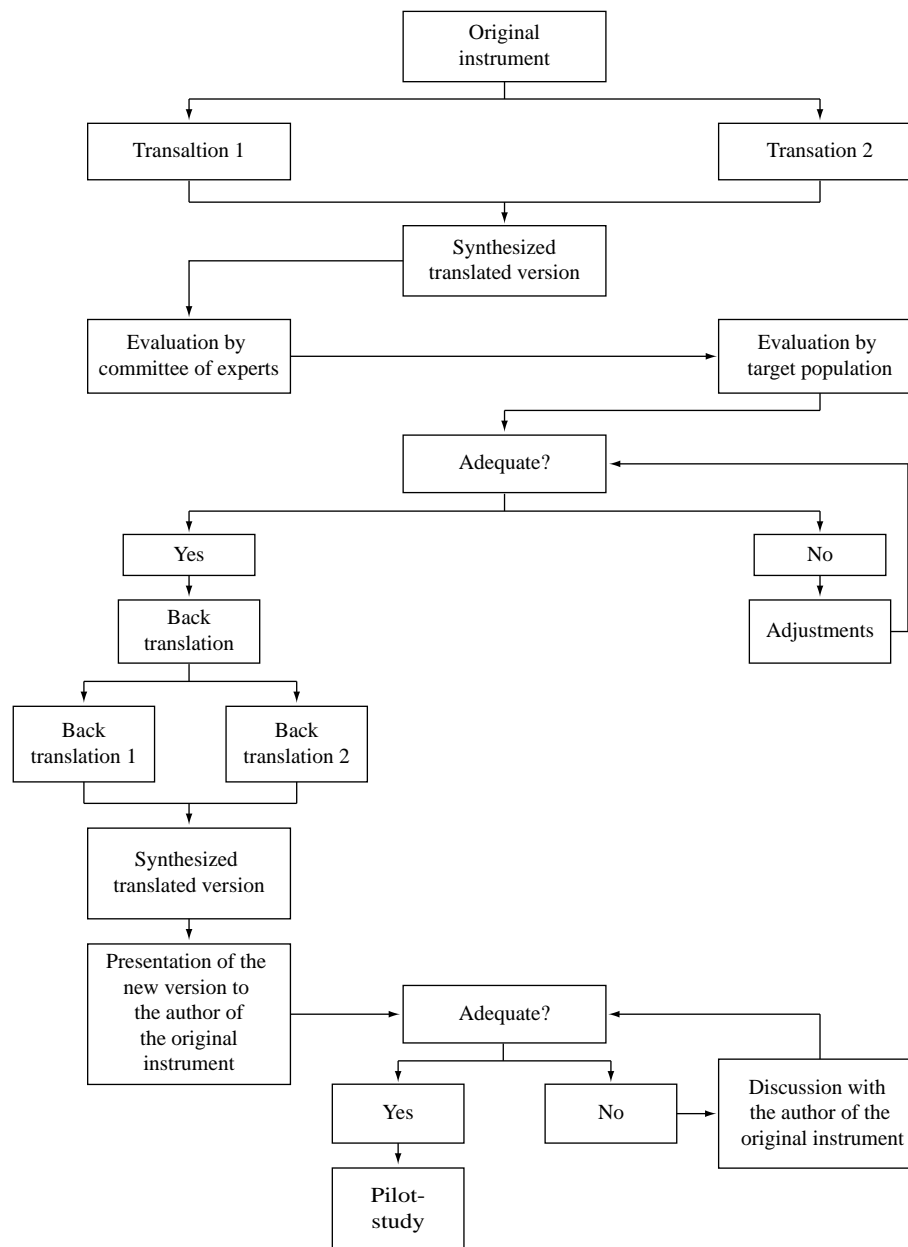


Figure 1. Procedures for cross-cultural adaptation of psychological instruments.

In some situations, there may be changes in the proposed steps, both between and within them. For example, there could be small pilot studies preceding back-translation, conduction of focus groups, or even no assessment by the target population. Sometimes, the instruments are

quite simple, easy to understand, and do not require evaluation by the target population. The same occurs when the instruments include non-verbal items, that is, their completion does not require reading, in which case the concern is merely about the translation of application

instructions, which need not comply with all the steps proposed in this article.

Aspects of Validating the Adapted Instrument

The previously mentioned adaptation processes aim to yield instruments that are equivalent across different cultures. For some authors (Herdman, Fox-Rushby, & Badia, 1997; Hui & Triandis, 1985), conceptual and idiomatic equivalence is the first aspect that is attained through the adaptation process. While qualitative methods are essential for ensuring the appropriateness of the adaptation process, they provide no information on the psychometric properties of the new instrument (Eremenco, Cella, & Arnold, 2005). Accordingly, complementary to the stages of instrument adaptation, statistical analyses must be performed to assess the extent to which the instrument can be considered valid for use in its designated context. Adapting and validating an instrument are, therefore, different but complementary steps. In general, scientific journals require that publications in this area clearly define both the adaptation and validation procedures.

The steps required during the validation of a psychological instrument are diverse (Urbina, 2007), and there is no consensus on how much validity the instrument must possess for it to be considered valid. We suggest that more evidence is better because this tends to increase measurement reliability. Additionally, like Urbina (2007), we advocate that other researchers should also evaluate such evidence, which further increases the instrument's validity.

The step of searching for evidence of an instrument's validity can be subdivided into two main areas: the first regarding the instrument validation for the new context and the second regarding the instrument validation for cross-cultural studies (involving different versions of the same instrument). In this article, we will discuss these aspects separately.

Evidence of Instrument Validity in the New Context

The first step in the validation of an instrument includes the evaluation of its factorial structure. In general, instruments are designed to measure constructs so that even when latent (that is, non-observable), they should have a relatively organized structure. Burnout, for example, is an occupational syndrome consisting of three different dimensions: emotional exhaustion, depersonalization, and low achievement at work.

The Maslach Burnout Inventory (MBI) (Maslach & Jackson, 1981), which is considered the most widely used instrument for its type of measurement, assesses those three characteristics through three different factors. Structures that are relatively similar to the original proposal are expected in MBI validation studies for use in new contexts. Otherwise,

the instrument will show discrepancies that affect the understanding of the construct under evaluation.

One should discuss possible changes that occur during validation studies in light of quantitative and qualitative aspects. By doing so, one can understand the possible reasons for changes in the factorial structure of the instrument. It is important to note that certain changes are expected as a result of sampling characteristics, especially in complex instruments, which have a high number of items and factors.

The techniques of exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) should be used to assist the researcher in his or her choice of a structure that is most plausible for the sample. Both EFAs and CFAs try to group a large number of observed variables with a small number of factors (latent dimensions) that explain the set of observed variables (Brown, 2006). For further details on these procedures, interested researchers should consult reference texts aimed at assisting with the completion of EFAs (Costello & Osborne, 2005; Damásio, in press) and CFAs (Brown, 2006).

As explained above, evaluating the factorial structure of the instrument considers only one aspect of the validation study. Other evidence of validity should be gathered, including evaluating the content and criterion validity of the instrument through the comparison of its results with those obtained by other equivalent measures. Analyzing the internal consistency between items, evaluating the precision (reliability and dependability), in addition to evaluating the consistency of the measurement at different times (temporal stability), are ways of finding evidence of validity of the adapted instrument. These procedures are performed after evaluating the factorial structure of the instrument and will not be presented in this article. For further clarification of this topic, we suggest reading Urbina (2007). In this book, the author presents specific chapters addressing the reliability and validity of psychological instruments.

Validation of Instruments for Cross-Cultural Studies

Another aspect concerning the validation of psychological instruments concerns the adaptation of the measure for use in cross-cultural studies. From this perspective, some authors use the concept of equivalence that refers not only to the qualitative aspects of the adapted instrument but also to the non-biased measurements between the adapted instrument and its original source. In this way, all the results of cross-cultural studies reflect only the actual differences (or similarities) between groups and are not the product of adaptation flaws (Eremenco et al., 2005).

Although EFAs and CFAs are commonly used for construct validation of adapted instruments, when aiming to conduct cross-cultural studies and compare various groups among them, the researcher must simultaneously

assess the measure's compatibility within the various groups (Hambleton & Patsula, 1998; Reise, Widaman, & Pugh, 1993; Sireci, 2005). Through those comparative analyses, the researcher ensures that the measurement similarly evaluates the same construct in different populations, and thereby ensures the assumption of measurement invariance (Reise et al., 1993). Multi-group confirmatory factor analysis (MGCFA), differential item functioning (DIF) proposed by the Item Response Theory (IRT), and multidimensional scaling (MDS) may be valuable ways of assessing measurement invariance (Milfont & Fischer, 2010; Sireci, 2005).

In the MGCFA, the factorial structure of the instrument is stipulated in advance, and the researcher simultaneously assesses the equivalence of structural parameters in the various groups of interest (Brown, 2006). Among various aspects, we may find the following: (1) the equivalence of instrument structure (i.e., whether the same number of factors and items per factor remain equivalent in the different groups); (2) the equivalence of item factorial loads (i.e., whether the weight or significance of items in the factor are similar in the different groups); (3) the similarity of covariance of the latent variable(s) (i.e., whether items explain the same variability level of the construct in the different groups, and/or whether the covariance between the instrument factors are similar in the different groups); and (4) the equivalence of residues of the observable variables (i.e., whether measurement errors are similar in the different groups) (Brown, 2006; Byrne, 2010).

Equivalence assessment of the structure and parameters of the test, through the MGCFA, answers a few key issues, including one in particular: Is the factorial structure of a given instrument equal between groups (do the same items assess the same construct)? Do items comprising a given factor hold the same significance within the different subgroups or show differences that preclude the comparison of different samples? Does the instrument show items that are biased towards one subgroup in particular?

The use of the MGCFA has grown exponentially in international studies because the technique enables an assessment of invariance for both the instrument structure and the various test parameters. Researchers who are interested in understanding the MGCFA further may refer to Brown (2006) and Byrne (2010).

The IRT, with the help of DIF, similarly enables the assessment of item similarities of a given instrument for different groups (Sireci, 2005). Under the IRT, a test item shows the DIF when the item response function (IRF) is different for subjects from different groups with the same level of the latent trait (Andrade, Laros, & Gouveia, 2010). If subjects have the same level of latent trait (for example, the same level of social phobia) but show different IRFs (different probabilities of answers and, therefore, different item scores), this item may be biased and thus may show differential operation. Two strategies may be adopted in

DIF situations. The first involves eliminating items with DIF so that groups are comparable. In those cases, the instrument becomes different from the original instrument, as some items are no longer used. The second proposal is to equate the scores of subjects maintaining the items with DIF (Eremenco et al., 2005). In those cases, the items with DIF are considered differently in the groups to maintain the equivalence between scores.

The DIF techniques proposed by the IRT are particularly useful for assessing biases of specific items and are not fruitful for assessing the equivalence of factorial structures (Kankaraš & Moors, 2010). One should also consider that most IRT models assess one-dimensional measurements exclusively (i.e., single-factor instruments). In the event of multidimensional instruments, DIF analyses are performed separately for the dimensions because for each dimension, the subjects will have specific levels of latent traits (Millsap, 2010). Several key texts may be consulted for a better understanding of the DIF techniques, as proposed by the IRT. For example, in Pasquali (2007), the author introduces specific chapters (chapters 7 and 8) on the topic. A practical example of the use of DIF techniques to assess research biases in cross-cultural studies can be found in Peterson et al. (2003).

Finally, the MDS includes statistical techniques that also enable the comparison of different groups simultaneously. Unlike the MGCFA, the MDS has the advantage that the factorial structure of the instrument does not need to be stipulated in advance. The researcher can compute different settings, choose the configuration of interest (for example, the configuration that best represents the theoretical structure of the instrument), and assess whether the structure is suitable for different groups (Arciniega, González, Soares, Ciulli, & Giannini, 2009). Another key MDS feature is that no linear model is needed to derive the underlying data structure, which is similar to the IRT (Sireci, 2005).

The validity of the assumption of factorial invariance between groups is crucial for the development and adaptation of psychometric instruments and for the comparison of groups in cross-cultural studies. Unless rigorously tested, researchers cannot claim that the structure and parameters of a given instrument are similar in different populations. If the instrument measurements are not comparable between the different groups, any differences found in terms of group scores or patterns of correlations with external variables are most likely measurement errors and thus do not reflect the actual differences between groups (Tanzer, 2005).

Final Considerations

In psychology, there is a growing interest in cross-cultural studies, which have demanded greater concern about the quality and suitability of adapted and validated

instruments for use in different contexts (ITC, 2010). While recognizing the importance of adapting instruments to other cultures, researchers have indicated that most of the research in this field has been deemed invalid because of inadequate procedures for translating and adapting the instruments (Hambleton, 2005). Sometimes, the adaptations of psychological instruments are based on the mere translation of items into the new language. In general, these translations are performed by the researchers themselves and rely solely on the process of back-translation, in which only the degree of semantic equivalence between the adapted version and the original version is analyzed (Cassepp-Borges et al., 2010; Hambleton, 2005; Reichenheim & Moraes, 2003).

There is no consensus on how to adapt an instrument for use in another cultural context. Such a procedure will depend on the instrument characteristics, the context of its application (both the original version and its adaptation), and the population for whom it is intended. The consensus is that the adaptation process, however, goes beyond mere translation, which does not guarantee construct validity or measurement reliability.

The process of adapting instruments should consider the relevance of original instrument concepts and domains in the new culture, in addition to considering the appropriateness of each item of the original instrument in terms of the ability to represent such concepts and domains in the new target population. Furthermore, the process should consider the semantic, linguistic, and contextual equivalence between the original and translated items and should include an analysis of the psychometric properties of the original instrument and its new version (ITC, 2010). Our experience in following these steps proposed has generated more reliable possibilities for evaluating various constructs in different contexts, while not wasting time, money, or materials. Poorly adapted instruments may present problems when they are used in other studies, which may generate inconsistent or unreliable data. In general, the researcher only realizes the errors of the process of translation, adaptation, and validation of an instrument at the time of collection and subsequent data analysis.

In cross-cultural studies, the use of instruments that are merely translated does not ensure reliable results because mere translation does not provide parameters to evaluate whether the results refer to differences or similarities between the different samples or derive from translation errors (Maneesriwongul & Dixon, 2004). During the last few decades, cross-cultural studies have attracted special attention from researchers, particularly in the field of mental health. These studies enable, through the application of a given instrument, comparisons between different individuals from different cultural contexts. Cross-cultural studies not only verify differences between individuals and cultures but also help us understand the common features among them. Therefore, we must have instruments that are

properly adapted and can provide measurement equivalence regardless of the context in which they are used. In this sense, besides the need for a rigorous process of adaptation, the assessment of the psychometric properties of the new instrument is essential to ensure that the instrument is in usable condition.

The present study introduces some procedures to be included in the adaptation process, in addition to statistical analyses that ensure that the instrument shows the necessary properties to be used both in the target population and in cross-cultural studies. Itemizing such aspects, specifically statistical procedures, is outside the scope of this article; however, our guidelines and references may serve as a basis for researchers to seek further developments in the subject of the adaptation of psychological instruments.

References

- Andrade, J. M., Laros, J. A., & Gouveia, V. V. (2010). O uso da teoria de resposta ao item em avaliações educacionais: Diretrizes para pesquisadores. *Avaliação Psicológica, 9*(3), 421-435.
- Arciniega, L. M., González, L., Soares, V., Ciulli, S., Giannini, M. (2009). Cross-cultural validation of the Work Values Scale (EVAT) using multi-group confirmatory factor analysis and confirmatory multidimensional scaling. *The Spanish Journal of Psychology, 12*(2), 767-772.
- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*(24), 3186-3191.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York: Routledge, Taylor & Francis.
- Cassepp-Borges, V., Balbinotti, M. A. A., & Teodoro, M. L. M. (2010). Tradução e validação de conteúdo: Uma proposta para a adaptação de instrumentos. In L. Pasquali, *Instrumentação psicológica: Fundamentos e práticas* (pp. 506-520). Porto Alegre: Artmed.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-9.
- Damásio, B. F. (no prelo). O uso da análise fatorial exploratória em psicologia. *Avaliação Psicológica*.
- Eremenco, S. L., Cella, D., & Arnold, B. J. (2005). A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Evaluation & the Health Professions, 28*(2), 212-232. doi:10.1177/0163278705275342

- 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
- Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology*, *61*(2), 29-45. doi:10.1027/1901-2276.61.2.29
- 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. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, *10*(3), 229-244.
- Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum.
- Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research*, *45*(1-3), 153-171. doi:10.1023/A:1006941729637
- Herdman, M., Fox-Rushby, J., & Badia, X. (1997). Equivalence and the translation and adaptation of health-related quality of life questionnaires. *Quality of Life Research*, *6*(3), 237-247. doi:10.1023/A:1026410721664
- Hui, C. H., & Triandis, H. C. (1985). Measurement in cross-cultural psychology: A review and comparison of strategies. *Journal of Cross-Cultural Psychology*, *16*(2), 131-152. doi:10.1177/0022002185016002001
- International Test Commission. (2010). *International Test Commission guidelines for translating and adapting tests*. Recuperado em 24 julho 2012, de <http://www.intestcom.org/upload/sitefiles/40.pdf>
- Kankaraš, M., & Moors, G. (2010). Researching measurement equivalence in cross-cultural studies. *Psihologija*, *43*(2), 121-136. doi:10.2298/PSI1002121K
- Maneesriwongul, W., & Dixon, J. K. (2004). Instrument translation process: A methods review. *Journal of Advanced Nursing*, *48*(2), 175-186. doi:10.1111/j.1365-2648.2004.03185.x
- Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Occupational Behavior*, *2*(2), 99-113. doi:10.1002/job.4030020205
- Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, *3*(1), 111-121.
- Millsap, R. E. (2010). Testing measurement invariance using item response theory in longitudinal data: An introduction. *Child Development Perspectives*, *4*(1), 5-9. doi:10.1111/j.1750-8606.2009.00109.x
- Oliveira, S. E. S., & Bandeira, D. R. (2011). Linguistic and cultural adaptation of the Inventory of Personality Organization (IPO) for the Brazilian culture. *Journal of Depression & Anxiety*, *1*(1), 1-7.
- Panzini, R. G., & Bandeira, D. R. (2005). Escala de coping religioso-espiritual (Escala CRE): Elaboração e validação de construto. *Psicologia em Estudo*, *10*(3), 507-516. doi:10.1590/S1413-73722005000300019
- Pasquali, L. (2007). *TRI - Teoria de Resposta ao Item: Teoria, procedimentos e aplicações*. Brasília: LabPAM.
- Peterson, M. A., Groenvold, M., Bjorner, J. B., Aaronson, N., Conroy, T., Cull, A., Fayers P., Hjermstad, M., Sprangers, M., Sullivan, M., & European Organization for Research and Treatment of Cancer Quality of Life Group. (2003). Use of differential item functioning analysis to assess the equivalence of translations of a questionnaire. *Quality of Life Research*, *12*(4), 373-385.
- Reichenheim, M. E., & Moraes, C. L. (2003). Adaptação transcultural do instrumento Parent-Child Conflict Tactics Scale (CTSPC) utilizado para identificar a violência contra a criança. *Cadernos de Saúde Pública*, *19*(6), 1701-1712. doi:10.1590/S0102-311X2003000600014
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*(3), 552-566. doi:10.1037/0033-2909.114.3.552
- Sireci, S. G. (2005). Using bilinguals to evaluate the comparability of difference language versions of a test. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 117-138). Mahwah, NJ: Lawrence Erlbaum.
- Sireci, S. G., Yang, Y., Harter, J., & Ehrlich, E. J. (2006). Evaluating guidelines for test adaptations: A methodological analysis of translation quality. *Journal of Cross-Cultural Psychology*, *37*(5), 557-567. doi:10.1177/0022022106290478
- Tanzer, N. K. (2005). Developing tests for use in multiple languages and cultures: A plea for simultaneous development. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 235-264). Mahwah, NJ: Lawrence Erlbaum.
- Urbina, S. (2007). *Fundamentos da testagem psicológica*. Porto Alegre: Artmed.
- Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage.
- Vivas, E. (1999). Estudios transculturales: Una perspectiva desde los trastornos alimentarios. In S. M. Wechsler & R. S. L. Guzzo (Orgs.), *Avaliação psicológica: Perspectiva internacional* (2a ed., pp. 463-481). São Paulo: Casa do Psicólogo.

Juliane Callegaro Borsa is Ph.D in Psychology of the Post-Graduate Program at the Universidade Federal do Rio Grande do Sul.

Bruno Figueiredo Damásio is Ph.D student of the Post-Graduate Program at the Universidade Federal do Rio Grande do Sul.

Denise Ruschel Bandeira is Associate Professor of the Post-Graduate Program at the Universidade Federal do Rio Grande do Sul.

Received: Jul. 24th 2012

1st revision: Sep. 19th 2012

Approved: Oct. 2nd 2012

How to cite this article

Borsa, J. C., Damásio, B. F., & Bandeira, D. R. (2012). Cross-cultural adaptation and validation of psychological instruments: Some considerations. *Paidéia (Ribeirão Preto)*, 22(53), 423-432. doi:<http://dx.doi.org/10.1590/1982-43272253201314>