



Methodology for cross-cultural adaptation of psychometric instruments

Wanderson Roberto Silva*

Universidade Estadual Paulista - UNESP, Araraquara, São Paulo, Brazil

wandersonroberto22@gmail.com

Juliana Dias

Universidade Estadual Paulista - UNESP, Araraquara, São Paulo, Brazil

julianacrdias@yahoo.com.br

David Costa

Instituto Universitário de Ciências Psicológicas, Sociais e da Vida - ISPA, Lisbon, Portugal

david_natanael@hotmail.com

Filipa Pimenta

Instituto Universitário de Ciências Psicológicas, Sociais e da Vida - ISPA, Lisbon, Portugal

filipa_pimenta@ispa.pt

João Maroco

ISPA- IU, Instituto Universitário de Ciências Psicológicas, Sociais e da Vida - ISPA, Lisbon, Portugal

jpmaroco@ispa.pt

Juliana Campos

Universidade Estadual Paulista - UNESP, Araraquara, São Paulo, Brazil

jucampos@fcfar.unesp.br

Abstract

Cross-cultural research is important to compare data from subjects with distinct cultural characteristics. The adaptation of psychometric measurement instruments is commonly observed in the literature. This study presents a methodology to perform the cross-cultural adaptation of instruments in different cultures using as example the Weight Concerns Scale (WCS) when applied to a sample of Brazilian and Portuguese female university students. Participated of this study 263 Portuguese students (mean=20.6 years, standard deviation = 1.8) and 252 Brazilian students (mean=21.1 years, standard deviation = 2.8). The participants responded the European and Brazilian Portuguese conciliated version of WCS. The face and content validity of the WCS were evaluated. Confirmatory factorial analysis was used and the goodness-of-fit indices applied were the chi-square by degrees of freedom ratio (χ^2/df), comparative fit index (CFI), normed fit index (NFI) and the root mean square error of approximation (RMSEA). The convergent validity was evaluated by the average variance extracted (AVE). The reliability was estimated by the composite reliability (CR) and internal consistency (α) of the measures. The factorial stability was checked by multiple-group analysis with groups' differences evaluated by $\Delta\chi^2$ tests. The conciliated version of the WCS presented adequate fit to total sample ($\chi^2/df = 6.57$; CFI= .97; NFI= .96; RMSEA= .10), Brazilian sample ($\chi^2/df= 5.37$; CFI= .95; NFI= .93; RMSEA=.13), and Portuguese sample ($\chi^2/df= 3.20$; CFI= .98; NFI= .97; RMSEA= .09). The WCS presented strong invariance between the two countries ($p > .05$). Analysis showed that the cross-cultural version adapted of WCS shows adequate validity, reliability and stability for both countries with different cultural contexts.

Keywords: cross-cultural; validation; reliability; methodology.

1. Introduction

Cross-cultural research has developed largely due to the need to compare data obtained from samples/populations with distinct cultural characteristics (Brislin, 1976). Maroco, Campos, Vinagre and Ribeiro (2014) report that the studies that do the cross-cultural adaptation of instruments are

important for the development of broad investigative methods that can be used with adequate validity and reliability in different populations with different socio-cultural contexts. This process is especially important when using psychometric scales.

He and Van de Vijver (2012) report that cross-cultural studies go beyond a simple process of translation and application of an instrument, in a given sample. These authors point out that the cross-cultural research should consider the fit of the model to different samples, to allow the collection of valid and reliable information. This process ensures adequate predictive ability of the instrument and enables the comparison of results obtained in different populations.

Still, it's important to remember that, before performing the translation and adaptation of instruments, researchers need to dominate the theory underlying the construction of the instrument (e.g. number of factors, their items, and correlations) and statistical modeling techniques necessary to evaluate the adequacy of this theory for different samples.

The translation and adaptation of instruments should consider a theory defined a priori and must follow a clear methodology and wherever possible should seek to draw up a conciliated version between countries of the same language (e.g. Portuguese from Brazil and Portugal). After completing the translation and adaptation of the conciliated version is necessary to test and evaluate their metric characteristics for independent samples. Generally, this evaluation is performed using confirmatory factor analysis followed by multi-groups analysis to verify the stability of the model in different samples (Hair et al., 2005; Kline, 1998; Maroco, 2014).

Thus, for conducting cross-cultural studies it is necessary that the researchers present a study design, and sample design compatible with all stages of psychometric evaluation of culturally adapted instruments (face validity, content and construct validities, stability and reliability) (Hair et al., 2005; Campos & Maroco, 2012; He & Van de Vijver, 2012).

Therefore, there is the growing interest of scientific research for carrying out cross-cultural studies since these studies can provide indications of interference of socio-cultural factors and comparison of different realities. The cross-cultural adaptation of a consolidated version of a measuring instrument can provide a valid and reliable instrument for use in different populations, and can thus increase the representation scientific work (Gato, Fontaine, & Leme, 2014). However, some cross-cultural studies presented in the literature do not show all steps to evaluate the psychometric properties of the instrument being validated. Thus, this study aimed to present a methodology, which uses confirmatory factor analysis to conduct a cross-cultural study in different cultures but with the same language.

2. Method

Below are the recommended steps for implementation of cross-cultural studies of psychometric instruments (Guillemin, Bombardier, & Beaton, 1993; Beaton et al., 2000; Hair et al., 2005; Maroco, 2014).

Face validity

The instrument should be submitted to face validation process:

- ✓ Translation and back translation: 3 independent and bilingual translators/back-translators.
- ✓ Development of a conciliated version (orthographic agreement: between regions and countries).
- ✓ Analysis of idiomatic, semantic, cultural and conceptual aspects - multidisciplinary team.
- ✓ Assessment of the incomprehension index (II) of all items: a pilot study - sample with the same characteristics of the final study: $II \leq 20\%$ (Guillemin, Bombardier, & Beaton, 1993; Beaton et al., 2000; Bradley, 1994).

Content validity

The adapted cross-cultural version should be submitted to an evaluation of the content validity (Laewshe, 1975).

- ✓ First, 5 to 40 experts in psychometric instruments and in the evaluated concept should judge each item of the instrument according to its essentiality (assessing it according to three options: essential, useful but not essential and not necessary).
- ✓ Thereafter the Content Validity Ratio (CVR) should be computed by Equation 1:

$$CVR = \frac{n - \left(\frac{N}{2}\right)}{\frac{N}{2}} \quad (1)$$

CVR: content validity ratio

n: number of experts indicated that the item is "essential"

N: total number of experts

- ✓ For the decision on the significance of each item, it will be used the recommendation by Wilson, Pan and Schumsky (2012), considering a 5% significance level.

Definitive study

Minimum sample size: 5 to 10 subjects per parameter of the model (Hair et al., 2005)

Psychometric Analysis

Item's sensitivity: skewness <|3| and kurtosis <|7| (Kline, 1998; Maroco, 2014).

Construct Validity

Construct validity should be assessed through the factorial validity, convergent and discriminant validities.

Confirmatory Factor Analysis

- ✓ Estimation methods: Maximum likelihood, non-weighted least squares, generalized least squares.
- ✓ Factor weight: $\lambda \geq .50$.
- ✓ Goodness-of-fit indices: absolute ($\chi^2/df \leq 2.0$; NFI $\geq .80$), relative (CFI and TLI $\geq .90$), parsimony (PGFI; PNFI and PCFI $\geq .60$) and discrepancy population (RMSEA $\leq .10$) (Byrne, 2001; Maroco, 2014).

Convergent validity

- ✓ Assessed by the Average Variance Extracted (AVE): AVE $\geq .50$ (Fornell & Larcker, 1981; Hair et al., 2005; Maroco, 2014).

Discriminant validity

- ✓ Used only when the instrument have at least two factors.
- ✓ AVE_i e $AVE_j \geq \rho_{ij}^2$.

Reliability

- ✓ Internal consistency: standardized Cronbach's alpha Coefficient (α): $\alpha \geq .70$ (Cronbach, 1951).
- ✓ Composite Reliability: CR $\geq .70$ (Fornell & Larcker, 1981).

Factorial Invariance

To check the stability of the factor structure of the instrument, a cross-validation between independent samples from two or more countries is required. Therefore, the total sample (all countries) should be randomly divided in "test sample" and "validation sample." Cross-validation will be performed in order to compare the rates observed in the test sample with another independent sample

arising from the same population. The invariance test can be done using factorial weights (λ) of two groups by multiple-group analysis using the chi-square statistical difference ($\Delta\chi^2$). When the hypothesis of the factorial weight's invariance is accepted (metric invariance), then it should be also done an analysis of the covariance's (scalar invariance) and residue's invariance (strong invariance). The same procedure should be carried out to verify the stability of transnational models, i.e. according to each country (Kaplan, 2000; Hair et al., 2005; Maroco, 2014).

Application of instrument

The methods presented above will be applied to the Weight Concerns Scale (WCS) whose Portuguese version reconciled (Brazil vs. Portugal) was presented by Dias, Silva, Maroco and Campos (2015). This version was produced by two Brazilian and two Portuguese researchers specialized in the area of body image/psychometrics to create, independently, a conciliated version of the instrument. Thereafter, this conciliated version was analyzed by Psychology Professionals and Portuguese teachers (two Brazilian and two Portuguese) to evaluate the idiomatic, semantic, cultural and conceptual equivalence of the instrument. The final conciliated version was pre-tested in a sample of 30 female university students (15 from each nationality) to estimate the incomprehension index (II) for each of the items. All items presented $II < .20$.

The WCS, originally in English, was proposed by Killen et al. (1994) to evaluate the body weight concern in the female population. The instrument was proposed as one-factor model including five questions, evaluated in a 5 to 7-point Likert-type response scale.

Despite the proposal by Dias, Silva, Maroco and Campos (2015), the cross-cultural stability of WCS was not evaluated. Thus, in the results section will be presented the psychometric characteristics of WCS, comparing Brazil vs. Portugal cross-cultural models. The WCS was applied to Brazilian ($n=252$) and Portuguese ($n=263$) female university students.

3. Results

Design Study: observational cross-sectional study

Sample Characterization: female university students – Course Area: Human and Social Sciences

Average age: Brazilian: 21.1 ± 2.8 year; Portuguese: 20.6 ± 1.8 ;

II: $< .20$

CVR: all items with $CVR_{8,.05} > .69$

Psychometric sensitivity: all items presented adequate values of the skewness and kurtosis.

Factor weights: $\lambda \geq .50$

Psychometric Properties: Table 1; Figure 1

Table 1. Confirmatory factor analysis (CFA) goodness of fit indices, Average variance extracted (AVE) and Reliability (Composite Reliability - CR, α) of the Weight Concerns Scale when applied to Portuguese and Brazilian university students

CFA								
Sample	n	χ^2/df	CFI	NFI	RMSEA	AVE	CR	α
Total	515	6.57	.97	.96	.10	.41	.77	.71
Brazilian	252	5.37	.95	.93	.13	.42	.78	.73
Portuguese	263	3.20	.98	.97	.09	.45	.79	.72

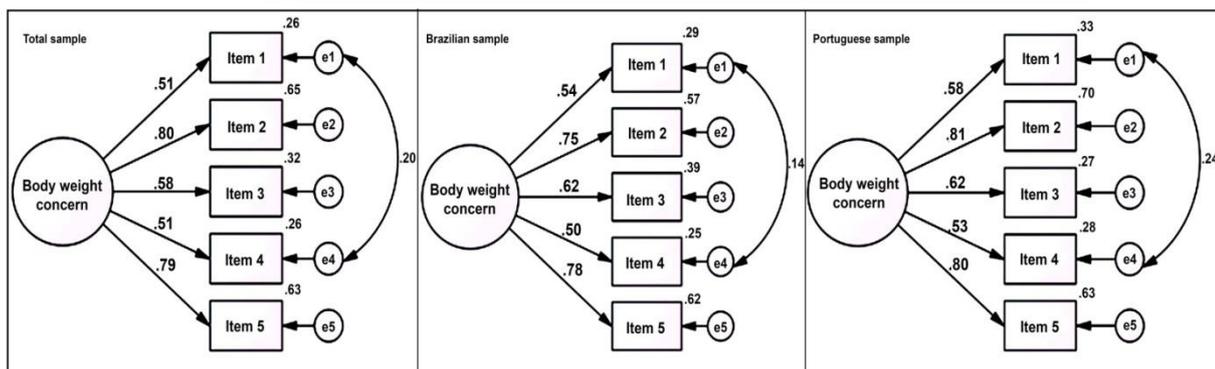


Figure 1. Factor Model of the Portuguese version of the Weight Concerns Scale applied to Portuguese and Brazilian university students

The invariance test showed strong invariance on two independent samples (test vs. validation) ($\Delta\chi^2_{\lambda}(4)=.44$, $p=.98$; $\Delta\chi^2_{Cov}(5)=.50$, $p=.99$; $\Delta\chi^2_{Res}(11)=9.96$, $p=.58$) and strong cross-cultural invariance (Brazil vs. Portugal) ($\Delta\chi^2_{\lambda}(4)=4.48$, $p=.30$; $\Delta\chi^2_{Cov}(5)=5.45$, $p=.36$; $\Delta\chi^2_{Res}(11)=13.24$, $p=.28$).

4. Discussion

The importance of this study relies on the detailed description of all steps necessary to perform the cross-cultural adaptation process of psychometric instruments and their exemplification. This work also sought to emphasize that the process of cross-cultural adaptation involves not only the steps of face validity and content, but also the application of the instrument in large samples to verify the psychometric properties of the adapted instrument, in independent samples and in different contexts.

These steps are essential to estimate the validity, reliability and stability of the adapted instrument. Also, before testing the factorial model's fit, it is crucial to develop a fully adapted version of the instrument, applicable in both countries/cultures (i.e. using language and content, equally and clearly understood in both countries/cultures). For example, English (e.g. USA vs. England) and Portuguese (e.g. Brazil vs. Portugal) are two languages which present cultural and language/content differences regarding their use in different countries. Therefore, after the development of a version of an instrument in a language used in different countries, it is necessary to check the cultural and content equivalence of the terms so that this version may be used in both countries, producing valid and comparable data.

In the example presented in the results' section we can see that the conciliated version in Portuguese of WCS (Dias, Silva, Maroco, & Campos, 2015) was adequately understood by university students. Regarding the content validity, the experts considered all WCS' items as essential to capture the construct of body weight concern. Then, we performed the evaluation of metric properties of the WCS, (using both the Brazilian and the Portuguese samples), to estimate the most adequate factorial validity and reliability for the total sample (Table 1, Figure 1). In this context it is important to note that the validity and reliability are not properties of the instruments per se, but dependent on the sample used (Campos, Bonafé, Dovigo, & Maroco, 2013). Thus, whenever the sample/population is different from the originally used in the validation process, new estimates to determine the psychometric characteristics of the instrument are mandatory (these estimates are required to be able to assess the quality of the information).

Furthermore, it is noteworthy that the convergent validity (AVE) is directly related to the factor weights of items. Due to this fact, and considering the example shown, the AVE could have been inappropriate because the discrepancy between the factor weights of each item in the body weight concern ($\lambda=.50-.80$). Regarding the chi-square statistic, it is observed the violation of cutoff point established; however, this fact generally occurs in studies with a large sample size since it is necessary to triple the minimum sample size to evaluate the stability of the factor structure of an instrument (Hair et al., 2005; Maroco, 2014).

Still, after the evaluation of the validity and reliability of the instrument, the stability of the factor structure should be investigated. Only after performing all these steps we can say that the cross-cultural adapted version is suitable for use.

5. Conclusion

To conduct cross-cultural adaptations of psychometric instruments it is required the application of a robust and careful methodology for the establishment of a culturally adapted version.

In addition, the metric properties (validity, reliability and stability) of the adapted version, evaluated in independent samples and different cultures, need to be estimated.

6. References

- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. Vol. 25, pp.3186-91.
- Bradley, C. (1994). Translation of questionnaires for use in different languages and cultures. In C. Bradley (Ed.), *Handbook of psychology and diabetes*. Chur, Switzerland: Harwood Academic Press. pp. 43-55.
- Byrne, M. (2001). *Structural equation modeling with Amos: basic concepts, applications and programming*. New Jersey: Lawrence Erlbaum Associates.
- Brislin, R. W. (1976). Comparative research methodology: Cross-cultural studies. *International Journal of Psych*. Vol.11, pp. 215–229.
- Campos, J. A. D. B., Bonafé, F. S. S., Dovigo, L. N., & Maroco, J. (2013). Avaliação psicométrica da escala de atitudes em relação à estatística. *Revista Brasileira de Biometria*. Vol.31, n.2, pp. 327-337.
- Campos, J. A. D. B., & Maroco, J. (2012). Maslach Burnout Inventory - Student Survey: Portugal-Brazil cross-cultural adaptation. *Revista Saúde Pública*. Vol.46, n.5, pp. 816-824.
- Cronbach, L. J. Coefficient alpha and the internal structure of tests. (1951). *Psychometrika*. Vol. 16, n.3, pp. 297-334.
- Dias, J. C., Silva, W. R., Maroco, J., & Campos, J. A. D. B. (2015). Validation of the Weight Concerns Scale in Brazilian university students. *Body Image*. In review
- Fornell, C., & Larcker, D. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*. Vol.18, n.1, pp.39-50.
- Gato, J., Fontaine, A. M., & Leme, V. B. R. (2014). Validação e Adaptação Transcultural da Escala Multidimensional de Atitudes Face a Lésbicas e a Gays. *Psicologia Reflexão e Crítica*. Vol.27, n.2, pp. 257-271.
- Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal Clinic Epidemiologic*. Vol.46, n.12, pp. 1417-32.
- Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2005). *Multivariate data analysis (6th ed.)*: Prentice Hall.
- He, J.; & Van de Vijver, F. (2012). Bias and equivalence in crosscultural research. *Online Readings in Psychology and Culture*. Retrieved from <http://scholarworks.gvsu.edu/orpc/vol2/iss2/8>.
- Kaplan, D. (2000). *Structural equation modeling: foundations and extensions*. Thousand Oaks, California: Sage Publications.
- Killen, J. D., Taylor, C. B., Hayward, C., Wilson, D. M., Haydel, K. F., Hammer, L. D., & Kraemer, H. (1994). Pursuit of thinness and onset of eating disorder symptoms in a community sample of adolescent girls: a threeyear prospective analysis. *International Journal of Eating Disorders*. Vol.16, n.3, pp.227-238.
- Kline R. B. *Principles and practice of structural equation modeling*. (1998). New York: The Guilford Press. pp.354.
- Laewshe CH. (1975). A quantitative approach to content validity. *Personnel Psychology*. Vol.28, n.13.
- Maroco, J., Campos, J. A. D. B., Vinagre, M. G., & Pais-Ribeiro, J. L. (2014). Adaptação Transcultural Brasil-Portugal da Escala de Satisfação com o Suporte Social para Estudantes do Ensino Superior. *Psicologia Reflexão e Crítica*. Vol.27, n.2, pp. 247-256.
- Maroco, J. (2014). *Análise de equações estruturais (2ª Edição)*. Pêro Pinheiro: ReportNumber.
- Wilson, F. R., Pan, W., & Schumsky, D. A. (2012). Recalculation of the critical values for Lawshe's content validity ratio. *Measurement and Evaluation in Counseling and Development*. Vol.45, pp.197-210.