

Developing Guidelines for Adapting Questionnaires into the Same Language in Another Culture

Desarrollo de Guías para Adaptar Cuestionarios Dentro de una Misma Lengua en Otra Cultura

Pablo Vallejo-Medina
Mayra Gómez-Lugo
Laurent Marchal-Bertrand
Alejandro Saavedra-Roa

SexLab KL, Fundación Universitaria Konrad Lorenz, Bogotá D.C., Colombia

Franklin Soler

Medicine and Health Sciences Faculty, Universidad del Rosario, Bogotá D.C., Colombia

Alexandra Morales

Health Psychology Department, Universidad Miguel Hernández de Elche, Alicante, Spain

Rec (10 de febrero de 2017) Accept (20 de junio de 2017)

Abstract

The adaptation of a test from a language into that same language in another culture is common; however, there are no clear guidelines for this process. The objective was to adapt a protocol providing some guidelines for adapting questionnaires from one language to the same language. A total of eight experts supported the adaption process and 825 participants from Spain and Colombia were evaluated in this study. Participants answered the brief version of the Sexual Assertiveness Scale, the Sexual Opinion Survey, the Massachusetts General Hospital-Sexual Functioning Questionnaire and the Sexuality Scale. The adaptation was made following some guidelines which have already been published. The results showed a strong partial invariance between countries. DIF analysis also replied this partial invariant form and adequate psychometric properties; guidelines to adapt questionnaires into same language in other cultures are presented. Therefore, the adaptation process - in the absence of further evidence - could be effective.

Keywords: Sexual assertiveness, psychometry, guidelines, adaptation, same language scales.

Resumen

La adaptación de test dentro de una misma lengua en varias culturas diferentes es común; sin embargo, no existen guías claras para realizar este proceso. El objetivo fue adaptar un protocolo generando unas guías para adaptar cuestionarios dentro de una misma lengua. Un total de ocho expertos realizaron el proceso de adaptación y 825 participantes de España y Colombia fueron evaluados en este estudio. Todos ellos contestaron a la versión breve de la *Sexual Assertiveness Scale*, la *Sexual Opinion Survey*, la *Massachusetts General Hospital-Sexual Functioning Questionnaire* y la *Sexuality Scale*. La adaptación se realizó siguiendo las directrices de algunas guías ya publicadas. Los resultados mostraron una invarianza fuerte entre los dos países. Estos hallazgos fueron replicados mediante DIF, además se observaron adecuadas propiedades psicométricas, finalmente las guías para el proceso de adaptación han sido presentadas. Por lo que concluimos que el proceso de adaptación – en ausencia de más evidencia- podría ser efectivo.

Palabras clave: asertividad sexual, psicometría, guía, adaptación, escalas mismo idioma.

* **Correspondence:** Pablo Vallejo Medina: Cra 9 Bis N° 62-43 Bogotá, Colombia pablo.vallejom@konradlorenz.edu.co

Agradecimientos: This work has been funded by Fundación Universitaria Konrad Lorenz, Acta 2014-003 95109141 granted to the corresponding author.

Introduction

The use of self-reports in psychology and related sciences and their adapting, translating and validation processes is quite common. Self-reports (tests, questionnaires, scales ...) are relevant when proving theories, making decisions about the effectiveness of a psychological treatment, experimentally verifying the impact of independent variables, etc. (Carretero-Dios & Pérez, 2007). Thus, the scores obtained from the scales have implications of great importance for the final result of any research that uses self-reports, as well as for the applied consequences deriving from the professional activity and decisions made based on these tests results (Padilla, Gomez, Hidalgo, & Muñoz, 2007). While there are a number of limitations inherent in these evaluation procedures in psychology (limitations in memory, desirability, understanding ...) there are also various standards to minimize said limitations. Following the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014), or the remarks of Kolen and Brennan (2014) and Downing and Haladyna's (2006) when preparing a self-report should minimize errors in the estimation of the construct to be evaluated. This should also improve the reliability of the evaluation. If the test does already exist in another language, these standards should be kept in mind, but an additional process should be implemented in order to adapt and translate such self-report so that the translation-related bias is reduced as far as possible.

Protocols and advice on test translation and adaptation have also been crucial, although they have received less attention than general standards. Currently, these adaptations are usually from English into another language i.e. Chinese, Spanish, German *inter alia* (Guillemin, Bombardier, & Beaton, 1993). In this case, Elosua, Mujika, Almeida and Hermosilla (2014), Hambleton, Merenda, and Spielberger (2005) or Muñoz and Bartram (2007) are some references to take into consideration when translating a test from one language into another. The aforesaid authors include explanations on translation techniques (back translation is usually recommended), cultural equivalence, or specification tables on content equivalence.

Recently, Muñoz, Elosua & Hambleton (2013) improved the traditional Guidelines for Adapting Tests (International Test Commission [ITC]; Hambleton, 1994) in a review of the standards (Hambleton, 1996) where some of their limitations are updated. They further propose - among other changes - replacing the back translation by an expert-agreed double forward translation. This effort seeks

to solve the problems posed by increasing transnational collaborations (Bullinger et al., 1998).

Spanish is one of the languages questionnaires are most often adapted to. This language is widely used (414 million native speakers; i.e. more than English. See Lewis, Gary, & Charles, 2014). These speakers are distributed in 33 countries (Lewis et al., 2014) and have large cultural differences. Although it is rare to find adaptations from American English to Australian or British English in English-speaking contexts like in Spanish, these do exist (McDowell, Courtney, Edwards & Shortridge-Baggett, 2005; Sanson-Fisher & Perkins, 1998). An example of these are the PISA test, where translation and adaptation processes are conducted following standardized guidelines (Programme for International Student Assessment, 2010). Nevertheless, it is not clear how they do that process in populations who have the same languages but different expressions and contexts, which undoubtedly may have a serious penalization in the data obtained (Grisay, 2003, 2007; Grisey & Monseur, 2007). Because there are no protocols (to the extent of our knowledge) to adapt self-reports from one source language into the same language in another culture, researchers are thus faced with a true challenge (McDowell et al., 2005).

Furthermore, evaluation in sexuality is usually modulated by moral, religious or ethical criteria from the context in which it is evaluated. In addition, intimate aspects of the person are inquired, thus making desirability higher than in other scales. For this reason, it was decided to test this protocol by using a sexuality scale (Sexual Assertiveness Scale; SAS Morokoff et al., 1997), considering that there are differences between the Colombian and Spanish cultures. This is a key variable in psychosexual wellbeing (Santos Iglesias & Sierra, 2010).

Consequently, the objective of this instrumental research is to adapt the protocol proposed by Muñoz et al. (2013), based on ITC guidelines (Hambleton 1994) in order to obtain new ones for adapting questionnaires from one source language to the same language in two different contexts. To this end, the procedure has been tested by using actual data answered on the SAS.

Method

Participants

Sample included a total of 825 participants who gave their acceptance to participate in this study by an informed

Table 1. *Sociodemographic Characteristics of the Sample and Differences between Countries*

Variables	Colombia	Spain	Contrast
Sex			
Female	282	219	$\chi^2 (1) = .48; p = .48$
Male	170	146	
Age	33.24(11.03)	35.60(12.94)	$t(823)=2.83; p < .01; d = 0.19$
Years of schooling	16.75(2.94)	15.64 (4.26)	$t(817)=4.42; p < .01; d = 0.30$
Sexual orientation			
Excl. heterosexual	392	315	$\chi^2 (7) = 4.71; p = .69$
2	28	24	
3	5	2	
4	4	5	
5	0	2	
6	6	8	
Excl. homosexual	15	13	
Asexual	2	1	
Marital status			
Single	198	182	$\chi^2 (3) = 22.39; p < .01; \eta^2 = 0.10$
Married	143	142	
Common-law marriage	86	29	
Divorced	23	16	
Religiousness			
Daily.	3	1	$\chi^2 (4) = 130.45; p < .01; \eta^2 = 0.38$
Once a week.	77	11	
Rarely in a month.	104	16	
Rarely in a year.	146	140	
Never.	117	200	

consent document. From this sample, 454 participants were Colombian and 371 participants were Spanish. All participants had had a stable partner for a minimum period of 6 months. Information was collected in over 100 different cities, in both countries. Other information about the sample is organized descriptively - and inferentially, if applicable - in Table 1.

Instruments

Sexual Assertiveness Scale (Morokoff et al., 1997). This study was based on the Spanish version of the scale, which has shown adequate psychometric properties (Sierra, Santos-Iglesias and Vallejo-Medina, 2012; Sierra, Vallejo-Medina and Santos-Iglesias, 2011; Vallejo-Medina and Sierra, 2015). This is composed by 18 items which evaluate three different dimensions: Initiation: ability to initiate

sexual relations when it is wanted, and to perform them as desired; refusal: ability to reject sexual unwanted practices or contact; and Sexually Transmitted Diseases - unwanted Pregnancy (STD-P). From these items, the nine items that were written backwards were eliminated for this study upon recommendation of the experts due to lack understanding of themselves. The instrument is available in both versions in Appendix A. High scores mean high sexual assertiveness.

Sexual Opinion Survey (SOS; Fisher, White, Byrne, & Kelley, 1988). The brief versions validated in Spain (Vallejo-Medina, Granados, & Sierra, 2014) and Colombia (Vallejo-Medina et al., 2016) were used. This six-item scale is answered on a six - alternative Likert scale, and evaluates attitudes toward sexuality, with a score ranging on a continuum between erotophobia and erotophilia. Reliability

of this survey in this study was .85. High scores indicate erotophilia.

Massachusetts General Hospital-Sexual Functioning Questionnaire (MGH-SFQ; Labbate & Lare, 2001). Spain (Sierra, Vallejo-Medina, Santos-Iglesias y Fernández, 2012) and Colombia versions (Marchal-Bertrand et al., 2016) were used. This evaluates sexual interest, sexual arousal, orgasm, erection (for men only) and overall satisfaction in both genders. Reliability in this study was .90. Higher scores indicate a better sexual performance.

Sexuality Scale (Wiedemann & Allgeier, 1993; Soler et al., 2016). The abbreviated scale has been used in this study. This consists of 15 items answered on a Likert scale. They are grouped in three dimensions: Sexual self-esteem, sexual depression and sexual preoccupation. Reliability of the sub-scales is appropriate: the Cronbach α ranges from .85 to .93 and test-retest reliability has significant minimum correlations of .67. Meanwhile, the external validity of the scale has proven adequate in English-speaking contexts.

Procedure

Preliminary SAS adaptation of the items from Spain to Colombia was made by four psychologists who were born in Colombia, or who were Colombian residents; all of them had at least a master's degree, and had been studying and living in Spain for at least two years. Using a table of specifications of the items, the psychologists had to indicate whether the items were fully understood, sounded strange, or whether they could be stated otherwise. This task was performed individually; it was compiled by the researchers of this study using a color table where the items were identified. Any items which had not been modified were identified in green; items which had been modified by only one of the psychologists were yellow, and those which had more than one modification, were red. Once the status of the items was identified, two sexuality experts met with the adapters in order to review the status of the items following Muñoz et al.'s (2013) guidelines. When the item was green, it was reported that the four experts had agreed not to modify it; however, they would read it again to ensure that the item had been understood correctly. Items that had a discrepancy (marked in yellow) were read again and it was stated that a different wording was considered more appropriate by one of the adaptation experts. At that point, a small debate was conducted under the guidance of the sexuality experts, until all the participants agreed on a final wording (adapter and sexuality experts). Something similar happened when the

items were marked in red. Suggested modifications were presented and debated in order to obtain the best solution.

During this adaptation process, an error occurred when adapting item number 4 where some content of the item was missed. Thus original item is: "I refuse to let my partner touch my breasts if I don't want that, even if my partner insists" and in the Colombian adaptation the - "if I don't want that"- fragment was skipped. The problem was mitigated by modifying the wording in the middle of the evaluation. This error has undoubtedly affected the results of this research and it will be corrected in the future. So, to control a possible change of contents of an item suggested by the adapters, adaptation should be confirmed with a backward "translation"; not as an adaptive method itself, but to ensure the avoidance of any alteration in the content of the item. Thus, it is recommended that once experts and adapters have agreed on the new items, at least one expert who did not partake in the process of adaptation assesses the equivalence of the content between the first Spanish version and the adapted version.

Once there was an agreement on the adapted version, a total of four different psychometrics and / or sexuality experts evaluated the items' properties. The following characteristics were evaluated: *Representativeness* and *Ownership* of the item to the sexual assertiveness construct; *Understanding* the item in the Colombian version; a single *Interpretation* (no ambiguity); and item *Clarity* (how concise it is). To this end, a table of specifications of the items and the ICAiken program (Soto & Segovia, 2009) were used. The ICAiken program allows to obtain the confidence interval for the Aiken V (1985). Experts scored the property of each item in a range of 1 (*Nothing*) to 4 (*Very*).

The sampling was conducted similarly in both countries; it was incidental and the evaluation was performed on-line (from October 23, 2014 to February 24, 2015). Questionnaires were designed in Typeform© and distributed through personal and Facebook contacts. The average time to answer the survey was 13 minutes and 18 seconds.

Data Analysis

A cut point below .50 in the lower limit (CI = .95%) of the Aiken V has been considered as a criterion of item inadequacy (Merino Soto & Livia Segovia, 2009).

EQS 6.1. multivariate software was used to calculate Factorial Invariance (FI). This was evaluated with a forward, multistage procedure, under a Mean and Covariances Structures procedure (MACS) as recommended (Byrne, 2009). This procedure allows for strong evaluation of invariance, as

compared to the Covariance Structures Analysis (COVS), which only allows for weak evaluation of the FI (Meredith, 1993). Moment analysis and Maximum Likelihood - Robust method (ML, Robust) were used. The latter is a robust estimator when non-multivariate normality is observed. Forward FI will be performed in four steps:

1. Configural invariance (invariance will be evaluated without restrictions in the model);
2. Metric or weak invariance (the factorial weights will be restricted);
3. Strong invariance (the intercepts will be restricted); and
4. Strict invariance (the variances of errors will be restricted).

The overall fit indices used were the Root Mean Square Error Approximation (RMSEA) and its 90% confidence interval, as well as the Comparative Fit Index (CFI). Values lower than .08 for the RMSEA and higher than .95 for the CFI will be considered of good fit. In addition, the CFI will be the main indicator used to evaluate the FI; the fact that the CFI does not decrease by more than .01 compared to the previous model (Cheung & Rensvold, 2002) shall be regarded as evidence of invariance. Finally, the Akaike Information Criterion (AIC) will also be reported. A considerable increase in this indicator will be indicative of absence of FI.

The SPSS 20.0 was used in order to evaluate the presence of DIF by way of multinomial logistic regression. This technique allows to detect uniform and non-uniform DIF in polytomous items (Hidalgo & López-Pina, 2004). If the contribution of DIF is uniform by Mean Square Error (99%), we will have a uniform DIF; if the contribution of Model 3 is significant (99%), DIF is not uniform. Given the large sample size, it is expected to find meaningful relationships; thus, so the DIF report will be complemented by a measure of effect size, which will measure the magnitude of DIF – the Nagelkerke's ΔR^2 ($R^2 < 0.035$ = negligible DIF; $0.035 < R^2 < 0.070$ = DIF moderate; $R^2 > 0.070$ = DIF high; Jodoin & Gierl, 2001). A purification process will also be implemented in stages for DIF items that have shown moderate or high levels. To this end, we will conduct a new regression eliminating all items with DIF from the scale. Thus, we will be able to determine whether the presence of DIF was attenuating, aggravating or concealing up the presence of more DIF (Differential Item Functioning).

The other results were obtained with the use of SPSS 20.0.

Results

Table 2 shows that the items adapted to the Colombian version have adequate qualitative properties, including an excellent content validity. This will indicate that a proper adaptation process of the items was performed. It is clear that this process failed to show the error in item 4.

Table 3 shows how the other psychometric properties of the items are suitable, except for item 4 of the Colombian version. Total item correlations of all items are above .30, and the Cronbach's alpha is always lower than the alpha of the corresponding sub-scale if the item is removed, except for items 1 of both versions. SD's are observed which indicate proper distribution of scores.

Subsequently, Factorial Invariance was tested. Mardia's test had a value of 6.27 and 7.40. Table 4 shows a suitable fit for the configural model. Adequate fit indices can be seen in the model of weak invariance; however, the AIC decreases slightly and the ΔCFI is six thousandths above .01. Therefore, the level of weak invariance could not be regarded as accepted. At this point, and in accordance with the modification indexes (Lagrange multiple test), partial invariance without item 4 constriction, was conducted. With this change, both good fit indexes and invariance indicators were observed. Continuing the progression, we find strong invariance which obtains adequate fit indices and it has appropriate indicators of factorial invariance. Finally, strict invariance, which is the last level of invariance tested here, was also fulfilled, thus closing the process of invariance. In addition, standardized weights of scales and related R2 can be observed in Table 5. DIF analysis showed no bias in the functioning of items – except item 4 - measuring assertiveness in both countries (see Table 6).

Indicators of external validity are shown in Table 7. Correlations are low and significant for most cases. As observed, differences between correlations in Spain and Colombia are very similar. Guidelines for adapting questionnaires in the same language source inter-culturally are described at Table 8.

Discussion

This paper presents a proposal that can be a guide for translations of scales from one language into the same

Table 2. *Properties of the Items.*

		Exp1	Exp2	Exp3	Exp4	M	Aiken	% agreement	LL 95%	UL 95%	
Item 1	Representativeness	3	4	4	4	3.75	.91	100	.64	.98	
	Ownership	1	1	1	1						
	Understanding	3	4	4	4	3.75	.91			.64	.98
	Interpretation	4	4	4	4	4	1			.75	1
	Clarity	4	4	4	4	4	1			.75	1
Item 2	Representativeness	3	4	4	4	3.75	.91	100	.64	.98	
	Ownership	1	1	1	1						
	Understanding	3	4	3	4	3.5	.83			.55	.95
	Interpretation	3	4	4	4	3.75	.91			.64	.98
	Clarity	3	4	4	4	3.75	.91			.64	.98
Item 3	Representativeness	3	4	4	4	3.75	.91	100	.64	.98	
	Ownership	1	1	1	1						
	Understanding	3	4	4	4	3.75	.91			.64	.98
	Interpretation	4	4	4	4	4	1			.75	1
	Clarity	4	4	4	4	4	1			.75	1
Item 4	Representativeness	3	4	4	4	3.75	.91	100	.64	.98	
	Ownership	2	2	2	2						
	Understanding	3	4	3	4	3.5	.83			.55	.95
	Interpretation	4	4	4	4	4	1			.75	1
	Clarity	3	4	4	4	3.75	.91			.64	.98
Item 5	Representativeness	3	4	4	4	3.75	.91	100	.64	.98	
	Ownership	2	2	2	2						
	Understanding	3	4	4	4	3.75	.91			.64	.98
	Interpretation	4	4	4	4	4	1			.75	1
	Clarity	4	4	4	4	4	1			.75	1
Item 6	Representativeness	3	4	4	4	3.75	.91	100	.64	.98	
	Ownership	2	2	2	2						
	Understanding	3	4	4	4	3.75	.91			.64	.98
	Interpretation	4	4	4	4	4	1			.75	1
	Clarity	3	4	4	4	3.75	.91			.64	.98
Item 7	Representativeness	3	4	3	4	3.5	.83	100	.55	.95	
	Ownership	3	3	3	3						
	Understanding	3	4	4	4	3.75	.91			.64	.98
	Interpretation	4	4	4	4	4	1			.75	1
	Clarity	4	4	4	4	4	1			.75	1
Item 8	Representativeness	3	4	4	4	3.75	.91	100	.64	.98	
	Ownership	3	3	3	3						
	Understanding	4	4	4	4	4	1			.75	1
	Interpretation	4	4	4	4	4	1			.75	1
	Clarity	4	4	4	4	4	1			.75	1
Item 9	Representativeness	3	4	4	4	3.75	.91	100	.64	.98	
	Ownership	3	3	3	3						
	Understanding	4	4	4	4	4	1			.75	1
	Interpretation	4	4	4	4	4	1			.75	1
	Clarity	4	4	4	4	4	1			.75	1

Note: Exp: Expert; LL: Lower limit; UL: Upper limit.

Table 3. *Some Psychometric Properties of the Items for samples from Colombia and Spain.*

Country	Dimension	Item	<i>M</i>	<i>SD</i>	Asymetry	Skewness	r_{it}^c	α -ítem	α	<i>M</i>	<i>SD</i>			
Colombia	Initiation	SAS1	2.57	1.07	-.33	-.73	.36	.82	.72	6.88	3.24			
		SAS2	2.34	1.44	-.36	-1.23	.69	.44						
		SAS3	1.98	1.50	.00	-1.48	.63	.52						
	Refusal	SAS4	0.59	0.98	1.85	2.51	.26	.71						
		SAS5	1.56	1.58	.46	-1.41	.53	.34				.61	3.87	3.13
		SAS6	1.76	1.53	.29	-1.46	.53	.34						
	STD-P	SAS7	1.25	1.50	.86	-.80	.82	.84						
		SAS8	1.49	1.67	.58	-1.39	.79	.86				.89	3.75	4.23
		SAS9	1.05	1.50	1.14	-.33	.79	.86						
Spain	Initiation	SAS1	2.45	1.16	-.26	-1.16	.41	.83	.75	6.78	3.28			
		SAS2	2.32	1.34	-.20	-1.34	.70	.53						
		SAS3	2.01	1.48	.06	-1.51	.67	.57						
	Refusal	SAS4	1.37	1.48	.65	-1.09	.63	.81						
		SAS5	1.96	1.65	.09	-1.67	.71	.74				.83	5.22	4.02
		SAS6	1.92	1.53	.08	-1.56	.72	.73						
	STD-P	SAS7	1.48	1.65	.59	-1.37	.81	.87						
		SAS8	1.76	1.79	.28	-1.75	.82	.86				.90	4.49	4.72
		SAS9	1.35	1.72	.74	-1.28	.81	.87						

Note. *M*: Mean; *SD*: Standard Deviation; r_{it}^c : item total corrected correlation; α -ítem: Cronbach alpha if item is deleted; α : Cronbach alpha.

Table 4. *Fit Indices for the Different Models Tested.*

Level of invariance	S-B χ^2	df	<i>p</i>	AIC	RMSEA	CI (90%) RMSEA	CFI	Δ CFI
Configural invariance	80.88	42	.00	-3.11	.049	.032 - .065	.986	-
Weak invariance	135.02	51	.00	33.02	.065	.052 - .078	.970	-.016
Weak -item 4	97.28	50	.00	-2.71	.049	.034 - .064	.983	-.003
Strong invariance	174.17	59	.00	56.17	.071	.059 - .083	.982	-.001
Strict invariance	178.18	68	.00	42.18	.065	.053 - .076	.983	.001

Note. S-B χ^2 : Satorra-Bentler scaled chi-square; df: degree of freedom; Δ CFI: CFI increment; Weak -item 4: Item 4 was not constricted.

Table 5. Standardized Weights, Errors and Explained Variance of each Item in the Configural Model for Colombia and Spain.

Items	Weight (λ)	Error	R ²
Colombia			
SAS1	.41	.91	.17
SAS2	.88	.46	.78
SAS3	.77	.65	.59
SAS4	.31	.95	.10
SAS5	.74	.67	.55
SAS6	.74	.67	.55
SAS7	.85	.51	.73
SAS8	.89	.45	.79
SAS9	.82	.56	.68
Spain			
SAS1	.44	.89	.19
SAS2	.89	.45	.79
SAS3	.80	.59	.64
SAS4	.70	.71	.49
SAS5	.81	.57	.66
SAS6	.83	.55	.69
SAS7	.85	.51	.73
SAS8	.89	.45	.79
SAS9	.87	.49	.76

Table 6. Differential Item Functioning Analysis

	Stage 1							Stage 2					
	Country				Country			Country			Country		
	Model 2	Model 3			Model 2	Model 3		Model 2	Model 3		Model 2	Model 3	
Ítem	$\chi^2_{(1)}$	<i>p</i>	R ²	$\chi^2_{(1)}$	<i>p</i>	R ²	$\chi^2_{(1)}$	<i>p</i>	R ²	$\chi^2_{(1)}$	<i>p</i>	R ²	
Initiation	1	1.08	0.29	0.002	1.50	0.22	0.003						
	2	0.20	0.65	0.001	0.63	0.42	0.001						
	3	0.62	0.43	0.001	0.01	0.89	0.000						
Refusal	4	45.05	0.00	0.072	4.17	0.41	0.006	62.92	0.001	0.101	9.18	0.002	0.014
	5	5.37	0.20	0.009	8.41	0.00	0.014	2.94	0.08	0.005	1.63	0.2	0.003
	6	25.14	0.00	0.041	4.09	0.04	0.006	2.94	0.08	0.005	0.58	0.44	0.001
STD-P	7	0.02	0.87	0.001	3.44	0.06	0.005						
	8	0.25	0.61	0.001	4.77	0.29	0.008						
	9	0.41	0.52	0.001	5.68	0.02	0.010						

Note. Stage 1 = Initial regression; Stage 2 = Purified regression. Model 1 regression with DIF absence, Model 2 is a regression with a grouping variable (uniform DIF) and Model 3 is added an interaction between group and total scoring in the test (non-uniform DIF)

Initiation: stage 1: Model 1 = $\chi^2_{(1)} = 0,2; p = .64 R^2 = 0.01$

Refusal: stage 1: Model 1 = $\chi^2(1) = 24,46; p = .000 R^2 = 0.019$. Stage 2 Model 1 = $\chi^2(1) = 6,59; p = .10 R^2 = 0.011$

STD-P: Stage 1: Model 1 = $\chi^2_{(1)} = 8,18; p = .004 R^2 = 0.013$.

Table 7. *Indicators of External Validity*

Colombia	1	2	3	4	5	6	7	8
Assertiveness initiation (1)	1	.22**	.16**	.30**	-.33**	-.01	.16**	.08
Assertiveness refusal (2)	.14**	1	.22**	-.02	-.03	-.23**	.06	-.28**
Assertiveness STD-P (3)	.04	.23**	1	-.02	-.08	-.07	.06	-.03
Sexual self-esteem (4)	.27**	-.07	-.10*	1	-.51**	.05	.09	.23**
Sexual depression (5)	-.28**	.05	.02	-.50**	1	.14**	-.09	-.42**
Sexual preoccupation (6)	.06	-.15**	-.03	.03	.12*	1	.11*	.20**
Attitudes toward sexuality (7)	.28**	-.04	-.02	.13**	.01	.27**	1	.01
Sexual Functioning (8)	.25**	-.17**	.02	.36**	-.51**	.22**	.02	1

Note. * = $p < .05$; ** = $p < .01$

Table 8. *Summarized guidelines from one language into the same language in another culture.*

1. Having a properly validated version of the test in the target language.
2. Taking into account the standard guidelines for adaptations and validations throughout the whole process.
3. It is suggested that 4 experts (with the characteristics below) be brought in the process of forward cultural adaptation of the linguistic content:
 - a. At least a master's degree in a related field.
 - b. The experts must have the nationality of the target adaptation.
 - c. The experts must have lived at least 2 years in the source culture of the adaptation.
4. The experts will individually evaluate 3 characteristics of the items in the original adaptation vis-à-vis the target culture:
 - a. If they are understood.
 - b. If they sound strange.
 - c. If they could be stated otherwise.
5. The experts will suggest an alternative wording, if applicable.
6. Experts' output will be compiled by the researchers of this study, using a color table where the items were identified as follows:
 - a. Green: no modifications or item issues were reported by none of them.
 - b. Yellow: one expert suggested modifications or item issues.
 - c. Red: two or more experts suggested modifications or item issues
7. Two researchers will meet with the adapters in order to review the status of the items, following Muñoz et al. (2013) guidelines.
 - a. If the item is green, it will be reported that the four experts had agreed not to modify it; however, it must be read it again so as to ensure that the item was understood correctly.
 - b. Items that had a discrepancy (marked in yellow or red) must be read again and a small debate, guided by the researchers until all the meeting participants have agreed on a final wording (adapter and researchers).
8. It is recommended that once researchers and adapters have agreed on the new items, at least one expert who did not partake in the process of adaptation assesses the equivalence of the content between the original adaptation and the new adapted version.
9. Once there is an agreement on the adapted version, a total of four different psychometrics and/or field experts will evaluate the items' properties as suggested by other authors. The following characteristics should be evaluated:
 - a. *Representativeness*
 - b. *Ownership*
 - c. *Understanding*
 - d. *Interpretation*
 - e. *Clarity*

Note. *It is worth mentioning that these guidelines are a suggestion and do not replace the use of international standards in adapting tests.*

language - which is commonly done amongst Iberoamerican countries (Cova, Bustos, Rincón, Grandón, Saldivia, & Inostroza, 2016; Londoño, Peñate, & González, 2016; Moyano-Díaz, Páez, & Torres, 2016; Ruiz et al., 2016; Ruiz, Suárez-Falcón, & Riaño-Hernández, 2016). This proposal has shown good results in adapting the brief SAS to Colombia, based on the Spanish version. Even though there are aspects to be improved, we believe that this work could provide guidance for adaptations amongst other countries and topics.

This research has proposed a 9-point guide to adapt tests from one language to the same language in another culture, minimizing adaptation issues. When using an instrument, it is necessary to reduce its distortion vulnerability due to factors that may affect measuring, like language (Arafat, Chowdhury, Qusar, & Hafez, 2016). However, it has been taken into account when adapting and translating processes, measurement errors happen when an adapted instrument is applied on a different population who speaks same language (Grisay, 2003, 2007; Grisey & Monseur, 2007). Thus, not only language is to be taken into account, but also cultural differences and linguistic expressions that can be dealt with in any given context. Table 8 – the main result – provides a number of specific guidelines that can be used as supplementary guidance in these adaptation processes.

Present guidelines were proved with the Spanish adaptation of the SAS (i.e., Spanish from Spain to Spanish from Colombia). The process was successfully carried out, in observance of the appropriate psychometric indicators. Problems in item 4 of the Colombian version are consistent in all the analysis (DIF, Invariance and psychometric properties) which show the relevance in keeping up with the guidelines proposed in this research. Invariance analysis is only defendable if item 4 is not constricted. Despite the adaptation problem that was mitigated in the middle of the evaluation, overall psychometric properties of the scale are good with a consistence similar to that observed in the original version (Morokoff et al., 1997), the Spanish version (Sierra et al., 2011) or others (Santos-Iglesias & Sierra, 2010). In addition, there was also enough external validity of the SAS with its respective relations, as observed before (Haavio-Mannila, & Kontula, 1997; Hurlbert, Singh, Menendez, Fertel, Fernández, & Salgado, 2005; Ménard & Offman, 2009); Santos-Iglesias, & Sierra, 2010; Santos-Iglesias, Sierra & Vallejo-Medina, 2013). In turn, this would be an indicator of an adaptation process.

In spite of the fact that a complete methodology was used, the adaptation process was not exempt from any mistakes. Hence, despite having obtained adequate content,

construct and external validity, as well as adequate levels of reliability, the adaptation process was not completely accurate. For the avoidance of errors, it is recommended to include a validation of the adapted form following the nature of the backward translation. While this adaptation method could be flawed on a standalone basis, in this case it would help to ensure that the adapted form is true to the original version.

It is worth mentioning that this work is, to the extent known, the first study conducted with this objective. Accordingly, the limitations are great and despite the effort, the method will have to be replicated using other measuring instruments and in other cultures. Moreover, this kind of adaptation from Spanish to Spanish could only be performed using initial or adapted versions which have been made following appropriate guidelines; otherwise, the result would be just a good adaptation of a scale which has been poorly adapted.

Ethical Statement

This study was approved by Konrad Lorenz Institutional Ethics Committee with research project number 2014-003 n°95109141 and it was respectful on human research by informing all procedures, risks and benefits in a consent document which was showed to all participants.

Conflict of Interest Statement

The authors declare no conflict of interest having an absence of any commercial or financial relationships that could influence this research.

References

- AERA, APA, & NCME (2014). *Standards for Educational and Psychological Tests*. Washington DC: American Educational Research Association
- Aiken, L. R (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45, 131-142. doi: 10.1177/0013164485451012
- Arafat, S. Y., Chowdhury, H.R., Qusar, M.S., & Hafez, M.A. (2016). Cross Cultural Adaptation & Psychometric Validation of Research Instruments: a Methodological Review. *Journal of Behavioral Health*, 5, 129-136. doi:10.5455/jbh.20160615121755
- Bennett, R. E. (2006). Inexorable and Inevitable: The Continuing Story of Technology and Assessment. In Bartram, D. & Hambleton R.K. (Eds.), *Computer-Based Testing and the Internet: Issues and Advances* (pp. 201-217). Chichester, England: Wiley.
- Bullinger, M., Alonso, J., Apolone, G., Lepège, A., Sullivan, M., Wood-Dauphinee, S., ... Ware, J. E. Jr. (1998). Translating health status questionnaires and evaluating their quality: the IQOLA Project approach.

- Journal of Clinical Epidemiology*, 51, 913-923. doi: [http://dx.doi.org/10.1016/S0895-4356\(98\)00082-1](http://dx.doi.org/10.1016/S0895-4356(98)00082-1).
- Byrne, B. M. (2009). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). New York, NY: Routledge/Taylor and Francis.
- Carretero-Dios, H., & Pérez, C. (2007). Normas para el desarrollo y revisión de estudios instrumentales: consideraciones sobre la selección de tests en la investigación psicológica. *International Journal of Clinical and Health Psychology*, 7, 863-882.
- Cheung, G. W., & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9, 233-255.
- Cova, F., Bustos, C., Rincón, P., Grandón, P., Saldívia, S., & Inostroza, C. (2016). Inventario de Conductas Infantiles en preescolares: Propiedades psicométricas del Inventario de Conductas Infantiles (CBCL/1.5-5) y del Informe del Cuidador/Educador (C-TRF) en Preescolares Chilenos. *Terapia Psicológica*, 34, 191-198.
- Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, 43, 121-149. doi: 10.1177/0748175610373459
- Downing, S. M., & Haladyna, T. M. (Eds.). (2006). *Handbook of test development*. Mahwah, NJ: LEA.
- Elosua, P., Mujika, J., Almeida, L.S., & Hermosilla, D. (2014). Procedimientos analítico-rationales en la adaptación de tests. Adaptación al español de la batería de pruebas de razonamiento. *Revista Latinoamericana de Psicología*, 46, 117-126 doi: 10.1016/S0120-0534(14)70015-9
- Fisher, W. A., White, L. A., Byrne, D., & Kelley, K. (1988). Erotophobia-erotophilia as a dimension of personality. *Journal of Sex Research*, 25, 123-151. doi:10.1080/00224498809551448
- Grisay, A. (2003). Translation procedures in OECD/PISA 2000 international assessment. *Language Testing*, 20, 225-240. doi: 10.1191/0265532203lt254oa
- Grisay, A. (2007). Translation equivalence across PISA countries. *Journal of Applied Measurement*, 8, 249-266.
- Grisay, A. & Monseur, C. (2007). Measuring the equivalence of item difficulty in the various versions of an international test. *Studies in Educational Evaluation*, 33, 69-86. doi: 10.1016/j.stueduc.2007.01.006
- Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health related quality of life measures: literature review and proposed guidelines. *Journal of Clinical Epidemiology*, 46, 1417-1432. doi:[http://dx.doi.org/10.1016/0895-4356\(93\)90142-N](http://dx.doi.org/10.1016/0895-4356(93)90142-N)
- Haavio-Mannila, E., & Kontula, O. (1997). Correlates of increased sexual satisfaction. *Archives of Sexual Behavior*, 26, 399-419. doi: 10.1023/A:1024591318836
- Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, 10, 229-240.
- Hambleton, R. K. (1996). Adaptación de tests para su uso en diferentes idiomas y culturas: fuentes de error, posibles soluciones y directrices prácticas. In J. Muñiz (Ed.), *Psicometría* (pp. 207-238). Madrid: Universitat.
- Hambleton, R. K., Merenda, P., & Spielberger, C. (Eds.). (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Hillsdale, NJ: Lawrence Erlbaum Publishers.
- Hidalgo, M. D. & López-Pina, J. A. (2004). Differential Item Functioning detection and effect size: A comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measurement*, 64, 903-915. doi: 10.1177/0013164403261769
- Hurlbert, D. F., Singh, D., Menendez, D.A., Fertel, E.R., Fernández, F., & Salgado, C. (2005). The role of sexual functioning in the sexual desire adjustment and psychosocial adaptation of women with hypoactive sexual desire. *Canadian Journal of Human Sexuality*, 14, 15-30.
- Jodoin, M. G., & Gierl, M.J. (2001). Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education*, 14, 329-349. doi: 10.1207/S15324818AME1404_2
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices*. Springer Science & Business Media. Iowa, USA.
- Labbate, L.A., & Lare, S.B. (2001). Sexual dysfunction in male psychiatric outpatients: Validity of the Massachusetts General Hospital Sexual Functioning Questionnaire. *Psychotherapy and Psychosomatics*, 70, 221-225. doi:10.1159/000056257
- Lewis, M.P., Gary, F.S., & Charles, D.F. (eds.). 2014. *Ethnologue: Languages of the world, (17th ed.)*. Dallas, TX: SIL International. Online version: <http://www.ethnologue.com>.
- Londoño, C., Peñate, W., & González, M. (2016). Validación Diferencial y Discriminante del Cuestionario de Depresión para Hombres (CDH). *Terapia Psicológica*, 34, 129-142.
- Marchal-Bertrand, L., Espada, J.P., Morales, A., Gómez-Lugo, M., Soler, F. & Vallejo-Medina, P. (2016). Adaptation, validation and reliability of the Massachusetts General Hospital-Sexual Functioning Questionnaire in a Colombian sample and factorial equivalence with the Spanish version. *Revista Latinoamericana de Psicología*, 48, 88-97. doi: 10.1016/j.rlp.2016.01.001
- McDowell, J., Courtney, M., Edwards, H., & Shortridge-Baggett, L. (2005). Validation of the Australian/English version of the Diabetes Management Self-Efficacy Scale. *International Journal of Nursing Practice*, 11, 177-184. doi:10.1111/j.1440-172X.2005.00518.x
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543. doi: 10.1007/BF02294825
- Ménard, A. D., & Offman, A. (2009). The interrelationships between sexual self-esteem, sexual assertiveness and sexual satisfaction. *The Canadian Journal of Human Sexuality*, 18(1/2), 35-45.
- Merino Soto, C., & Livia Segovia, J. (2009). Intervalos de confianza asimétricos para el índice la validez de contenido: Un programa Visual Basic para la V de Aiken. *Anales de Psicología*, 25, 169-171.
- Morokoff, P.J., Quina, K., Harlow, L.L., Whitmire, L., Grimley, D.M., Gibson, P.R., & Burkholder, G.J. (1997). Sexual Assertiveness Scale (SAS) for women: development and validation. *Journal of Personality and Social Psychology*, 73, 790-804.
- Moyano-Díaz, E., Páez, D., & Torres, M. (2016). Propiedades psicométricas del cuestionario para medir estrategias de aumento de la felicidad (HIS) en versión castellana (CEA-EAP). *Terapia Psicológica*, 34, 143-154.
- Muñiz, J., & Bartram, D. (2007). Improving international tests and testing. *European Psychologist*, 12, 206-219. doi: <http://dx.doi.org/10.1027/1016-9040.12.3.206>
- Muñiz, J., Elosua, P., & Hambleton, R.K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*, 25, 151-157.
- Nunnally, J.C., Bernstein, I.H., Arellano, J.A.V., & Guillén, M.T. (1995). *Teoría psicométrica*. México: Mcgraw-hill.
- Padilla, J. L., Gómez, J., Hidalgo, M. D., & Muñiz, J. (2007). Esquema conceptual y procedimientos para analizar la validez de las consecuencias del uso de los test. *Psicothema*, 19, 173-178.
- Programme for International Student Assessment (2010). Translation and adaptation guidelines for PISA 2012 (Report No. NPM(1010)4e). Retrieved from: <https://www.oecd.org/pisa/pisaproducts/49273486.pdf>
- Ruiz, F. J., Suárez-Falcón, J. C., Barón-Rincón, D., Barrera-Acevedo, A., Martínez-Sánchez, A., & Pena, A. (2016). Factor structure and psychometric properties of the Dysfunctional Attitude Scale Revised in Colombian undergraduates. *Revista Latinoamericana de Psicología*, 48, 81-87. doi: 10.1016/j.rlp.2015.10.002
- Ruiz, F. J., Suárez-Falcón, J. C., & Riaño-Hernández, D. (2016). Psychometric properties of the Mindful Attention Awareness Scale in Colombian undergraduates. *Suma Psicológica*, 23, 18-24. doi: 10.1016/j.sumpsi.2016.02.003
- Sanson-Fisher, R.W., & Perkins, J.J. (1998). Adaptation and validation of the SF-36 Health Survey for use in Australia. *Journal of Clinical Epidemiology*, 51, 961-967. doi: [http://dx.doi.org/10.1016/S0895-4356\(98\)00087-0](http://dx.doi.org/10.1016/S0895-4356(98)00087-0)

- Santos-Iglesias, P., & Sierra, J. C. (2010). El papel de la asertividad sexual en la sexualidad humana: una revisión sistemática. *International Journal of Clinical and Health Psychology, 10*, 553-577.
- Santos-Iglesias, P., Sierra, J. C., & Vallejo-Medina, P. (2013). Predictors of sexual assertiveness: The role of sexual desire, arousal, attitudes, and partner abuse. *Archives of Sexual Behavior, 42*, 1043-1052. doi: 10.1007/s10508-012-9998-3
- Sierra, J. C., Vallejo-Medina, P., Santos-Iglesias, P., & Fernández, M. L. (2012). Validación del Massachusetts General Hospital-Sexual Functioning Questionnaire (MGH-SFQ) en población española. *Atención Primaria, 44*, 516-524.
- Sierra, J. C., Santos-Iglesias P., & Vallejo-Medina, P. (2012). Evaluación de la equivalencia factorial y métrica de la Sexual Assertiveness Scale (SAS) por sexo. *Psicothema, 24*, 316-322.
- Sierra, J. C., Vallejo-Medina, P., & Santos-Iglesias, P. (2011). Propiedades psicométricas de la versión española de la Sexual Assertiveness Scale (SAS). *Anales de Psicología, 27*, 17-26.
- Soler, F., Gómez-Lugo, M., Espada, J. P., Morales, A., Sierra, J. C., Marchal-Bertrand, L., & Vallejo-Medina, P. (2016). Adaptation and Validation of the Brief Sexuality Scale in Colombian and Spanish Populations. *International Journal of Psychology and Psychological Therapy, 16*, 343-356.
- Soto, C. M., & Segovia, J. L. (2009). Confidence intervals for the content validity: a visual basic com-puter program for the aiken's v. *Anales de Psicología, 25*, 169-171.
- Vallejo-Medina, P., & Sierra, J. C. (2015). Adaptation and validation of the Sexual Assertiveness Scale (SAS) in a sample of male drug users. *The Spanish Journal of Psychology, 18*, E21. doi: 10.1017/sjp.2015.25
- Vallejo-Medina, P., Granados, M. R., & Sierra, J. C. (2014). Propuesta y validación de una versión breve del Sexual Opinion Survey en población española. *Revista Internacional de Andrología, 12*, 47-54.
- Vallejo-Medina, P., Marchal-Bertrand, L., Gómez-Lugo, M., Espada, J. P., Sierra, J. C., Soler, F., & Morales, A. (2016). Adaptation and validation of the Brief Sexual Opinion Survey (SOS) in a Colombian sample and factorial equivalence with the Spanish version. *PloS one, 11*, e0162531. doi: <http://dx.doi.org/10.1371/journal.pone.0162531>
- Wiedeman M. W., & Allgeier E. R. (1993). The measurement of sexual-esteem: Investigation of Snell and Papini's (1989) Sexuality Scale. *Journal of Research in Personality, 27*, 88-102. <http://dx.doi.org/10.1006/jrpe.1993.1006>

Appendix A

This scale aims to evaluate the perception or appreciation that you have about your sexual behavior. Please check your level of agreement / disagreement regarding each question. In both versions (i.e. Colombian and Spanish versions) are placed the original English items in *italic* which were taken from Morokoff et al. (1997).

	Nunca / Never	A veces / Sometimes	La mitad de las veces / Half of the time	Casi siempre / Almost always	Siempre / Always
Colombian Sexual Assertiveness Scale <i>adaptation</i> (SAS; Morokoff et al., 1997)					
1. Inicio las relaciones sexuales con mi pareja cuando lo deseo. / <i>I begin sex with my partner if I want to.</i>	0	1	2	3	4
2. Le indico a mi pareja que me toque los genitales cuando así lo deseo. / <i>I let my partner know if I want my partner to touch my genitals.</i>	0	1	2	3	4
3. Le indico a mi pareja que me estimule los genitales con su boca cuando así lo deseo. / <i>I let my partner know if I want to have my genitals kissed.</i>	0	1	2	3	4
4. No dejo que mi pareja acaricie mi cuerpo si no quiero, incluso cuando insiste. / <i>I refuse to let my partner touch my breasts if I don't want that, even if my partner insists.</i>	0	1	2	3	4
5. Si he dicho que no, no dejo que mi pareja me toque los genitales aunque insista. / <i>If I said no, I won't let my partner touch my genitals even if my partner pressures me.</i>	0	1	2	3	4
6. Me niego a tener sexo si no me provoca, incluso si mi pareja insiste. / <i>I refuse to have sex if I don't want to, even if my partner insists.</i>	0	1	2	3	4
7. Cuando tengo relaciones sexuales con mi pareja me aseguro de utilizar condón. / <i>I make sure my partner and I use a condom or latex barrier when we have sex.</i>	0	1	2	3	4
8. Insisto en usar condón cuando quiero, aunque mi pareja prefiera no usarlo. / <i>I insist on using a condom or latex barrier if I want to, even if my partner doesn't like them.</i>	0	1	2	3	4
9. Me niego a tener relaciones sexuales si mi pareja no quiere utilizar el condón. / <i>I refuse to have sex if my partner refuses to use a condom or latex barrier.</i>	0	1	2	3	4

Spanish Sexual Assertiveness Scale *adaptation* (SAS; Morokoff et al., 1997)

	Nunca / Never	A veces / Sometimes	La mitad de las veces / Half of the time	Casi siempre / Almost always	Siempre / Always
1. Inicio las relaciones sexuales con mi pareja cuando lo deseo. / <i>I begin sex with my partner if I want to.</i>	0	1	2	3	4
2. Le indico a mi pareja que me toque los genitales cuando así lo deseo. / <i>I let my partner know if I want my partner to touch my genitals.</i>	0	1	2	3	4
3. Le indico a mi pareja que me estimule los genitales con su boca cuando así lo deseo. / <i>I let my partner know if I want to have my genitals kissed.</i>	0	1	2	3	4
4. Me niego a dejar que mi pareja acaricie mi cuerpo si no lo deseo, incluso cuando insiste. / <i>I refuse to let my partner touch my breasts if I don't want that, even if my partner insists.</i>	0	1	2	3	4
5. Si he dicho que no, no dejo que mi pareja me toque los genitales aunque me presione. / <i>If I said no, I won't let my partner touch my genitals even if my partner pressures me.</i>	0	1	2	3	4
6. Me niego a tener sexo si no me apetece, incluso si mi pareja insiste. / <i>I refuse to have sex if I don't want to, even if my partner insists.</i>	0	1	2	3	4
7. Cuando tengo relaciones sexuales con mi pareja me aseguro de utilizar condón o barrera de látex. / <i>I make sure my partner and I use a condom or latex barrier when we have sex.</i>	0	1	2	3	4
8. Insisto en usar condón o barrera de látex cuando quiero, incluso aunque mi pareja prefiera no usarlos. / <i>I insist on using a condom or latex barrier if I want to, even if my partner doesn't like them.</i>	0	1	2	3	4
9. Me niego a tener relaciones sexuales si mi pareja no quiere utilizar condón o barrera de látex. / <i>I refuse to have sex if my partner refuses to use a condom or latex barrier.</i>	0	1	2	3	4