



This is a repository copy of *What happens when individuals answer questionnaires in two different languages*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/175782/>

Version: Published Version

Article:

Paz, C., Hermosa-Bosano, C. and Evans, C. orcid.org/0000-0002-4197-4202 (2021) What happens when individuals answer questionnaires in two different languages. *Frontiers in Psychology*, 12. 688397. ISSN 1664-1078

<https://doi.org/10.3389/fpsyg.2021.688397>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



What Happens When Individuals Answer Questionnaires in Two Different Languages

Clara Paz^{1†}, Carlos Hermosa-Bosano¹ and Chris Evans^{2*†}

¹ Escuela de Psicología, Universidad de Las Américas, Quito, Ecuador, ² Department of Psychology, The University of Sheffield, Sheffield, United Kingdom

OPEN ACCESS

Edited by:

Simone Grassini,
Norwegian University of Science and
Technology, Norway

Reviewed by:

Shelia Kennison,
Oklahoma State University,
United States
Henrike Katina Blumenfeld,
San Diego State University,
United States

*Correspondence:

Clara Paz
clara.paz@udla.edu.ec
Chris Evans
chris@psyctc.org

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 30 March 2021

Accepted: 25 May 2021

Published: 23 June 2021

Citation:

Paz C, Hermosa-Bosano C and
Evans C (2021) What Happens When
Individuals Answer Questionnaires in
Two Different Languages.
Front. Psychol. 12:688397.
doi: 10.3389/fpsyg.2021.688397

The aim of the present study was to compare scores from the English and the Spanish versions of two well-known measures of psychological distress using a within-subject approach. This method involved bilingual participants completing both measures in four conditions. For two groups of people, measures were offered in the same language both times and for the other two groups, each language version was offered, the order differing between the groups. The measures were the Clinical Outcomes in Routine Evaluation-Outcome Measure and the Schwartz Outcome Scale-10, both originally created in English and then translated to Spanish. In total, 109 bilingual participants (69.7% women) completed the measures in two occasions and were randomly allocated to the four conditions (English-English, English-Spanish, Spanish-English and Spanish-Spanish). Linear mixed effects models were performed to provide a formal null hypothesis test of the effect of language, order of completion and their interaction for each measure. The results indicate that for the total score of the Clinical Outcomes in Routine Evaluation-Outcome Measure just language had a significant effect, but no significant effects were found for completion order or the language by order interaction. For the Schwartz Outcome Scale-10 scores, none of these effects were statistically significant. This method offers some clear advantages over the more prevalent psychometric methods of testing score comparability across measure translations.

Keywords: translation, CORE-OM, SOS-10, outcomes measures, psychological interventions, score comparability, cultural adaptation

INTRODUCTION

Translation of psychological measures is a common practice worldwide and there is steadily growing interest in comparing findings across countries and languages. More than 30 guidelines have been created to guide the translation process as well as the cultural adaptation of these measures; however, there is no consensus of which one is the best methodology (Epstein et al., 2015). These procedures all aim to attain equivalence between the original and the translated version of the instrument, but they vary in how much they acknowledge that perfect equivalence is an ideal that is not ensured by any translation method nor even easy to fully define. Despite these challenges, clearly, it is not possible to compare data across translations of measures without some empirical exploration of score comparability. Generally, this is explored using between-subjects approaches that compare the scores given to the measure by persons from different populations looking for

“measurement invariance,” a statistical property that indicates that the same construct is measured across samples (Vandenberg and Lance, 2000; Byrne and Watkins, 2003; Milfont and Fischer, 2010). Measurement invariance is usually tested within either Classical Test Theory (CTT), typically through Confirmatory Factor Analysis (CFA); or within Item Response Theory (IRT) methods. These psychometric approaches are sophisticated and, when their assumptions are met, they offer power to detect forms of non-equivalence. However, with two different models, these methods test the covariance of items *across the individuals* between the language samples. This is highly appropriate for measures largely designed to compare individuals’ scores at a single completion, e.g., to determine school or university entry or to measure personality traits. However, this is tangential to the aims of measures of within individual change, i.e., to the typical aim of measures used in psychotherapy to assess change over time (Tarescavage and Ben-Porath, 2014). For such measures these forms of measurement invariance are rarely found, even within one language (Kim et al., 2010; Fried et al., 2016). In addition, such measures cannot be used for single item measures such as visual analog scales as such measures have no item covariance to explore. This situation creates a need for other approaches to explore the equivalence of measures across translations.

The aim of the present study was to compare scores from the English and the Spanish versions of two well-known measures of psychological distress using a within-subject approach, a rarely used approach (Spector et al., 2015). One variant of this approach is to offer each language version of the questionnaire to the same group of individuals on two occasions, perhaps randomizing their order (Rivas-Vazquez et al., 2001; Chen and Bond, 2010; Chen et al., 2014; Zavala-Rojas, 2018; Rezapour and Zanjirani, 2020). However, another variant throws more light on language and order effects by using four groups, again with two occasions. For two of the four groups measures are offered in the same language on each occasion; for the other two groups, each language is offered, the order differing between groups. Such studies aiming to estimate any language effect are compromised by the test-retest effect: the very common finding that when mental health measures are completed twice in non-help-seeking samples there is often a mean shift between occasions (see Durham et al., 2002, for a review). The four-group design disaggregates the test-retest effect from any effect of language allowing that test-retest mean change might interact with language, something that cannot be done in the two group method. Studies using the two group method, specifically to test measures used to assess change in psychotherapy, are scarce. The only study that we found is the one conducted by Wiebe and Penley (2005) using the Beck Depression Inventory, and there, no significant language effects on mean scores were founded. However, studies using that approach for personality measures have found language effects for some traits (Chen and Bond, 2010; Chen et al., 2014; Rezapour and Zanjirani, 2020), suggesting that language might activate cognitive styles and behavioral expressions which are linked to the specific linguistic-social context in which the language was learned or in which it is most commonly used. In relation to time, the study conducted by Wiebe and Penley (2005) reported that the scores were lower

in the second completion for all groups, thus indicating that time produced an effect. This study reports the use of the four group method with the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM; Evans et al., 2000, 2002) and the Schwartz Outcome Scale-10 (SOS-10; Blais et al., 1999), both originally created in English and then translated to Spanish. Both measures are widely used in various languages to track outcomes and change in psychological distress when applying psychological interventions. Also, both measures can be used free of charge which contributes to their growing use in Latin American countries in recent years (Paz et al., 2020a,b). These factors led us to choose these measures to test our method and add to the literature about them in Latin America.

METHODS

Participants

Participants were bilingual (Spanish-English and English-Spanish) adults, living in Ecuador, a predominantly Spanish-speaking country, who have either completed an International Baccalaureate or have obtained an English proficiency certificate. Participants were recruited by means of posts on alumni social media pages of high schools which offer International Baccalaureate with English as the main language. Also, we asked institutions that offer English lessons to distribute the invitation to participate in the study to people who had attained an English proficiency certificate. Participation in the study was entirely voluntary with no monetary incentives or compensation for the participants.

In total, 167 persons completed the measures on the first occasion, 110 (65.9%) were women and 57 (34.1%) were men. The mean age was 26.41 ($SD = 7.78$) and the age range varied from 18 to 58. Of the 167 participants, 155 described themselves as bilingual and met the eligibility criteria, for four of them English was the native language. On the second occasion measures were completed by 109 participants, 70.3% of those who completed the measures on the first occasion met eligibility criteria. Of these, 76 (69.7%) were women and 33 (30.3%) were men. The mean age was 26.36 ($SD = 7.26$) with range from 18 to 50. No significant effects of gender [$\chi^2(1, N = 155) = 1.30, p = 0.253$] or age [$t(77) = 0.11, p = 0.91, d = 0.02$] were found comparing those who only completed the measures on the first occasion with those completing the measurements on both occasions. More fundamentally, the language in which the questionnaire was first presented was not statistically significantly related to non-completion on second occasion: $\chi^2(1, N = 155) = 3.31, p = 0.07$. The mean number of days between first and second occasion was 20.1 ($SD = 6.68$) ranging from 14 to 40 days. The breakdown into groups of those who participated on both occasions was English-English = 20, English-Spanish = 32, Spanish-English = 32 and Spanish-Spanish = 25.

Measures

CORE-OM

This instrument is a self-report questionnaire containing 34 items that assess general psychological state (Evans et al., 2000, 2002). The Spanish translation (Feixas et al., 2012) was conducted in

Spain (Trujillo et al., 2016) and their psychometric properties of the scores were good and similar to those reported for the original version. The psychometric properties of the Spanish version offered to an Ecuadorian sample (Paz et al., 2020b) indicated that these properties are similar to those reported for the scores of the original version in United Kingdom (Evans et al., 2002) and to the Spanish version in Spain. In the present study, the Cronbach alpha of the English version was 0.96, 95% CI [0.93, 0.98] and that of the Spanish version was 0.93, 95% CI [0.90, 0.95].

SOS-10

This instrument is a 10-item self-report measure of well-being (Blais et al., 1999). This measure was translated to Spanish in the United States with a group of bilingual individuals (Rivas-Vazquez et al., 2001), and the exploration of the psychometric properties of the scores obtained in that study indicate that they are good and comparable with those found for the original version in English (Blais et al., 1999). The psychometric properties of this measure have been also tested in Ecuador (Paz et al., 2020a). Results from this study indicate that the properties are similar to those found for the original and the Spanish translations (both conducted in United States). In the present study Cronbach alpha was 0.94, 95% CI [0.90, 0.96] for the English version and 0.93, 95% CI [0.89, 0.95] for the Spanish version.

Procedures

The sample size was calculated by simulation which showed that a sample size of at least 100 participants, assuming a minimal test-retest stability of 0.6 would give 95% effect size confidence intervals of ± 0.16 for the effect of language assuming equal sized groups.

Participants were contacted through social media and invited to participate in an online anonymous study conducted using LimeSurvey (LimeSurvey Project Team, 2012). Participants who gave informed consent completed a brief sociodemographic questionnaire and were then randomly allocated to complete the measures (CORE-OM and SOS-10) either in Spanish or in English. Then, 14 days later participants were invited to complete the measures again and they were randomly allocated again to complete the measures either in Spanish or in English. Hence four conditions of the presentation of the measures were created: (1) English on both occasions (EE), (2) English the first occasion and Spanish on the second occasion (ES), (3) Spanish the first occasion and English on the second occasion (SE), and (4) Spanish on both occasions (SS). If participants did not complete measures on the second occasion, they were reminded weekly, until the termination of data collection. The random allocation aimed to balance the groups, but it was recognized that random allocation and attrition would be very unlikely to achieve perfectly balanced group allocation. The survey was set up with all questions mandatory hence there were no missing data. The Ethics Committee of the Universidad de Las Américas, Ecuador approved the study [ID: 2020-0619].

Data Analysis

Linear mixed effects models were performed to provide a formal null hypothesis test of the effect of language, order

of completion and their interaction. Regression models were conducted separately for each measure (CORE-OM and SOS-10). As there can be gender differences in mean scores on such measures, gender was recorded and entered as a simple participant-level co-variable. However, statistically significant gender effects were not expected given the sample size, between groups test and the low expected effect. Age effects are generally very small, and the age range of the participant pool was small; for these reasons it was not treated as a covariate. These analyses were conducted using the nlme: Linear and Non-linear Mixed Effects Models package (Pinheiro et al., 2021) from R statistical software (R Core Team, 2020).

RESULTS

To test language effects, measurement completion order and their interaction, independent models were performed for each measure. Total scores of each measure were included as the dependent variables and gender, language, completion order and the interaction of language and order of completion were included as fixed effects, while the subjects were placed as random effects for each model.

For the total score of the CORE-OM, results indicated that neither of completion order ($\beta = 0.02$, $SE = 0.06$, $df = 106$, $p = 0.78$) nor participants' gender ($\beta = -0.11$, $SE = 0.13$, $df = 107$, $p = 0.36$) showed statistically significant effects. Language did show a significant effect ($\beta = -0.18$, $SE = 0.07$, $df = 106$, $p = 0.02$) with no significant interaction with order of completion ($\beta = 0.03$, $SE = 0.11$, $df = 106$, $p = 0.82$). **Figure 1** shows the violin plot of the CORE-OM total scores by language, gender and occasion. In this figure a tendency for total scores on the Spanish version to be higher than on the English version is visible, however the difference is within the precision of estimation of the means as shown by the vertical bootstrap 95% confidence interval lines crossing the shared mean scores for each occasion. Mean differences and effect sizes of the CORE-OM scores by group are presented in **Table 1**.

For the SOS-10 scores, no significant effects were found for completion order ($\beta = -0.51$, $SE = 1.32$, $df = 106$, $p = 0.69$), gender ($\beta = 0.07$, $SE = 2.45$, $df = 107$, $p = 0.97$), language ($\beta = 0.34$, $SE = 1.48$, $df = 106$, $p = 0.82$), nor for the interaction between the order of completion and the language ($\beta = 1.41$, $SE = 2.23$, $df = 106$, $p = 0.52$). **Figure 2** shows the violin plot of the SOS-10 total scores by language, gender and occasion which shows that scores for the English version tended to be higher than for the Spanish version, but the difference is clearly within the precision of estimation of the means. Mean differences and effect sizes of the SOS-10 scores by group are presented in **Table 1**.

DISCUSSION

This is one of the few studies that have empirically tested the effects of language and completion order of two psychological distress measures in a bilingual sample (Spanish and English). The results indicate that for the total score of the CORE-OM, language had a small but statistically significant effect, but neither gender, completion order, nor the language by completion order

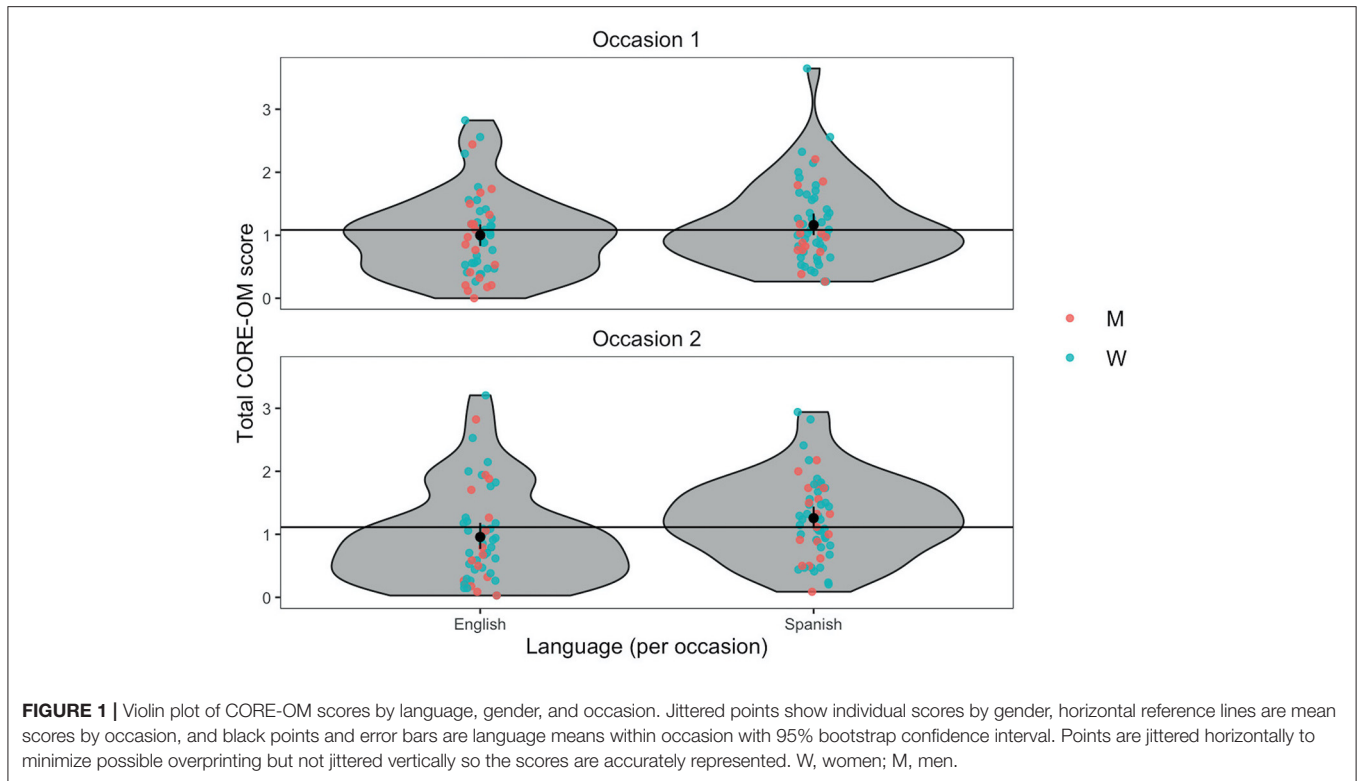


FIGURE 1 | Violin plot of CORE-OM scores by language, gender, and occasion. Jittered points show individual scores by gender, horizontal reference lines are mean scores by occasion, and black points and error bars are language means within occasion with 95% bootstrap confidence interval. Points are jittered horizontally to minimize possible overprinting but not jittered vertically so the scores are accurately represented. W, women; M, men.

TABLE 1 | Mean differences and effect sizes for the effect of time and language.

Variable	CORE-OM		SOS-10	
	Mean difference <i>M</i> [95% CI] ^a	Effect size Hedges' <i>g</i> [95% CI] ^a	Mean difference <i>M</i> [95% CI] ^a	Effect size Hedges' <i>g</i> [95% CI] ^a
English-English	-0.11 [-0.60, 0.32]	-0.13 [-0.80, 0.40]	1.25 [-7.65, 9.65]	0.09 [-0.53, 0.72]
English-Spanish	-0.17 [-0.45, 0.12]	-0.28 [-0.76, 0.21]	-0.47 [-6.44, 5.40]	-0.04 [-0.53, 0.45]
Spanish-English	0.15 [-0.16, 0.50]	0.22 [-0.24, 0.73]	-1.94 [-7.72, 3.50]	-0.16 [-0.67, 0.31]
Spanish-Spanish	-0.03 [-0.36, 0.30]	-0.04 [-0.64, 0.49]	1.40 [-4.92, 7.64]	0.12 [-0.40, 0.72]

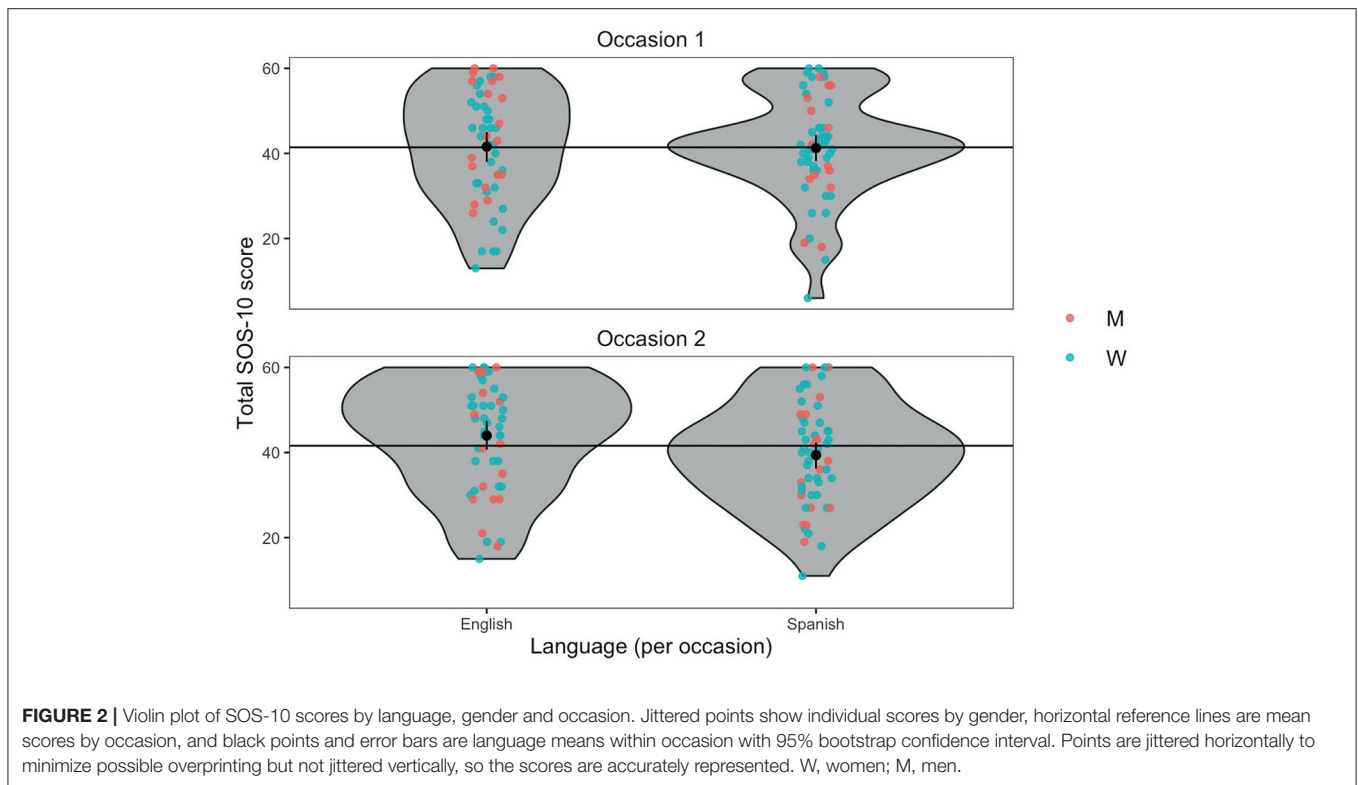
^aBootstrapped 95% confidence interval.

interaction presented significant effects. For the SOS-10 none of these effects were significant. Then, it seems that Spanish and English versions of the CORE-OM are not perfectly equivalent, but the differences, smaller than one point on a score with range from 0 to 40 and with 95% confidence intervals for the differences also under one score point, were sufficiently small and sufficiently precisely estimated to suggest that language change in a population speaking both English and Spanish is not likely to invalidate the use of change scores nor to rule out comparison of score changes from samples in either language. Detected language effect and possible explanations for their existence might be explored in future studies with this specific and other similar measures.

Language effects have been founded in previous studies using personality measures (Chen and Bond, 2010; Chen et al., 2014; Rezapour and Zanjirani, 2020), but not when using measures of psychological distress (Wiebe and Penley, 2005). The results of our study indicate the presence of this effect in one measure

but not in the other, which suggest that further studies have to be conducted with this type of measures to arrive to more consistent conclusions. However, there is also evidence that response styles are different in different languages and countries (Harzing, 2006), so that possibly explain the presence of a language effect in our data.

The present study has several limitations. First, this study was conducted completely online due the restrictions imposed by the Coronavirus Disease-19 pandemic during 2020; in Ecuador, paper-and-pencil format remains much commoner than online. However, the pandemic brings is probably changing this so generalizability may be less affected by this in the future. Second, in this study we only considered two occasions for assessment. A common practice in psychological intervention is to apply the same instrument at different moments of the intervention. Whether resources are available, future studies might test more complex designs, including more assessment occasions, and might seek bigger samples.



A more general limitation of this method is that people must have sufficient competence in both languages to participate. In countries where bilingualism is common this may not limit generalizability, however, in countries where bilingualism in the chosen languages is uncommon, as in Ecuador, such a method should be complemented by conventional explorations in unilingual samples. There is psychometric literature suggesting that language in use by bilingual or multilingual people can affect questionnaire responses, and variables such as age of acquisition of the languages, dominance and proficiency can affect reception of, and communicating ideational and emotional material in each language (Paradis, 2008; Pavlenko, 2008). Rather separate from that literature there is largely qualitative literature about the bilingualism in psychological therapies, *inter alia* de Zulueta (1995), Pérez Foster (1996), Dale and Altschuler (1997), Das (2020) and it may be relevant that most of the development of analytic therapies in the late 19th and 20th centuries was led by bilingual or multilingual individuals, often working in contexts in which unilingualism was rare. In our study most of the bilingual participants were using their second language (English for most) in academic contexts which can impact on socialization and integration of affective words (Pavlenko, 2008), words that are commonly present in therapy measures. Clearly we can never know if people with fluency in two languages will experience questionnaires in either language exactly as their peers lacking this fluency. A useful extension of this study might be to evaluate and categorize participants' levels of fluency in the languages which might be used as covariates in the four group

design and future studies might ask about proficiency, ages of acquisition and contexts where each language is used. Equally, a qualitative extension exploring the experiences when answering the measures could add to the quantitative findings.

Despite these limitations this study presents a new method to evaluate score comparability for translated mental health measures and we believe the method offers advantages over, and a very useful complement to the more prevalent between groups exploration of measurement invariance in terms of cross-sectional item score covariance. This approach maps more clearly than those methods to the widespread use of such measures to evaluation within-individual change rather than between-individual score comparisons (Newsom, 2015) and can be applied to single item measures.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study will be publicly available. The data will be found at figshare repository: The datasets presented in this study can be at: figshare (doi: 10.6084/m9.figshare.14336609).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Universidad

de Las Américas [ID: 2020-0619]. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CP and CE designed the study, performed data analysis, and reported results. CE conducted the power estimation simulations. CP and CH-B implemented the study and collected data. CP, CE, and CH-B participated in writing the article. All authors agree to be accountable for the content of the work.

REFERENCES

- Blais, M. A., Lenderking, W. R., Baer, L., DeLorell, A., Peets, K., Leahy, L., et al. (1999). Development and initial validation of a brief mental health outcome measure. *J. Pers. Assess.* 73, 359–373. doi: 10.1016/S1098-3015(11)70926-3
- Byrne, B. M., and Watkins, D. (2003). The issue of measurement invariance revisited. *J. Cross. Cult. Psychol.* 34, 155–175. doi: 10.1177/0022022102250225
- Chen, S. X., Benet-Martínez, V., and Ng, J. C. K. (2014). Does language affect personality perception? A functional approach to testing the whorfian hypothesis. *J. Pers.* 82, 130–143. doi: 10.1111/jopy.12040
- Chen, S. X., and Bond, M. H. (2010). Two languages, two personalities? Examining language effects on the expression of personality in a bilingual context. *Personal. Soc. Psychol. Bull.* 36, 1514–1528. doi: 10.1177/0146167210385360
- Dale, B., and Altschuler, J. (1997). “Different language/different gender: Narratives of inclusion and exclusion,” in *Multiple Voices. Narrative in Systemic Family Psychotherapy*, eds. K. Papadopoulos and J. Byng-Hall (London: Routledge), 125–145.
- Das, S. (2020). Multilingual matrix: exploring the process of language switching by family therapists working with multilingual families. *J. Fam. Ther.* 42, 39–53. doi: 10.1111/1467-6427.12249
- de Zulueta, F. (1995). Bilingualism, culture and identity. *Gr. Anal.* 28, 179–190. doi: 10.1177/0533316495282007
- Durham, C. J., McGrath, L. D., Burlingame, G. M., Schaalje, G. B., Lambert, M. J., and Davies, D. R. (2002). The effects of repeated administrations on self-report and parent-report scales. *J. Psychoeduc. Assess.* 20, 240–257. doi: 10.1177/07342829020200302
- Epstein, J., Santo, R. M., and Guillemin, F. (2015). A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *J. Clin. Epidemiol.* 68, 435–441. doi: 10.1016/j.jclinepi.2014.11.021
- Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., et al. (2002). Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *Br. J. Psychiatry* 180, 51–60. doi: 10.1192/bjp.180.1.51
- Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., Conell, J., and McGrath, G. (2000). CORE: clinical outcomes in routine evaluation. *J. Ment. Heal.* 9, 247–255. doi: 10.1080/jmh.9.3.247.255
- Feixas, G., Evans, C., Trujillo, A., Ángel, L., Botella, L., Corbella, S., et al. (2012). La versión española del CORE-OM : clinical outcomes in routine evaluation - outcome measure. *Rev. Psicoter.* 23, 109–135. doi: 10.33898/rdp.v23i89.641
- Fried, E. I., van Borkulo, C. D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., and Borsboom, D. (2016). Measuring depression over time... Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychol. Assess.* 28, 1354–1367. doi: 10.1037/pas0000275
- Harzing, A. W. (2006). Response styles in cross-national survey research: a 26-country study. *Int. J. Cross Cult. Manag.* 6, 243–266. doi: 10.1177/1470595806066332
- Kim, S.-H., Beretvas, N. S., and Sherry, A. R. (2010). A validation of the factor structure of OQ-45 scores using factor mixture modeling. *Meas. Eval. Couns. Dev.* 42, 275–295. doi: 10.1177/0748175609354616
- LimeSurvey Project Team (2012). *LimeSurvey: An Open Source Survey Tool*. Hamburg: LimeSurvey Project Team.
- Milfont, T. L., and Fischer, R. (2010). Testing measurement invariance across groups: applications in cross-cultural research. *Int. J. Psychol. Res.* 3, 111–130. doi: 10.21500/20112084.857
- Newsom, J. T. (2015). *Longitudinal Structural Equation Modeling: A Comprehensive Introduction*. New York, NY: Routledge
- Paradis, M. (2008). Bilingualism and neuropsychiatric disorders. *J. Neurolinguistics* 21, 199–230. doi: 10.1016/j.jneuroling.2007.09.002
- Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism* 11, 147–164. doi: 10.1017/S1366728908003283
- Paz, C., Evans, C., Valdiviezo-Oña, J., and Osejo-Taco, G. (2020a). Acceptability and psychometric properties of the Spanish translation of the Schwartz Outcome Scale-10 (SOS-10-E) outside the United States: a replication and extension in a Latin American context. *J. Pers. Assess.* 10, 1–10. doi: 10.1080/00223891.2020.1825963
- Paz, C., Mascialino, G., and Evans, C. (2020b). Exploration of the psychometric properties of the clinical outcomes in routine evaluation-outcome measure in Ecuador. *BMC Psychol.* 8:94. doi: 10.1186/s40359-020-00443-z
- Pérez Foster, R. (1996). The bilingual self duet in two voices. *Psychoanal. Dialogues* 6, 99–121. doi: 10.1080/10481889609539109
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and Team, R. C. (2021). *nlme: Linear and Nonlinear Mixed Effects Models*. Available online at: <https://cran.r-project.org/package=nlme> (accessed January 10, 2021).
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Available online at: <http://www.r-project.org/> (accessed January 10, 2021).
- Rezapour, R., and Zanjirani, S. (2020). Bilingualism and personality shifts : different personality traits in persian- english bilinguals shifting between two languages. *Iran. J. Learn. Memory* 3, 23–30. doi: 10.22034/iepa.2020.230347.1169
- Rivas-Vazquez, R. A., Rivas-Vazquez, A., Blais, M. A., Rey, G. J., Rivas-Vazquez, F., Jacobo, M., et al. (2001). Development of a Spanish version of the schwartz outcome scale-10: a brief mental health outcome measure. *J. Pers. Assess.* 77, 436–446. doi: 10.1207/S15327752JPA7703_05
- Spector, P. E., Liu, C., and Sanchez, J. I. (2015). Methodological and substantive issues in conducting multinational and cross-cultural research. *Annu. Rev. Organ. Psychol. Organ. Behav.* 2, 101–131. doi: 10.1146/annurev-orgpsych-032414-111310
- Tarescavage, A. M., and Ben-Porath, Y. S. (2014). Psychotherapeutic outcomes measures: a critical review for practitioners. *J. Clin. Psychol.* 70, 808–830. doi: 10.1002/jclp.22080
- Trujillo, A., Feixas, G., Bados, A., García-Grau, E., Salla, M., Medina, J. C., et al. (2016). Psychometric properties of the Spanish version of the clinical outcomes in routine evaluation - outcome measure. *Neuropsychiatr. Dis. Treat.* 12, 1457–1466. doi: 10.2147/NDT.S103079
- Vandenberg, R., and Lance, C. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices. *Organ. Res. Methods* 3, 4–70. doi: 10.1177/109442810031002
- Wiebe, J. S., and Penley, J. A. (2005). A psychometric comparison of the Beck depression inventory - II in English and Spanish. *Psychol. Assess.* 17, 481–485. doi: 10.1037/1040-3590.17.4.481

FUNDING

The present study was funded by Dirección General de Investigación, Universidad de Las Américas, Quito, Ecuador [PSI.CPE.20.04].

ACKNOWLEDGMENTS

We want to acknowledge Gabriel Osejo-Taco for his effort and support during data collection. In addition, we want to thank two reviewers for their inputs which substantially improved the paper.

Zavala-Rojas, D. (2018). Exploring language effects in crosscultural survey research: does the language of administration affect answers about politics? *Methods Surv. Methodol.* 12, 127–150. doi: 10.12758/mda.2017.16

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Paz, Hermosa-Bosano and Evans. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.