



A Spanish Translation of the Hoarding Rating Scale: Differential Item Functioning and Convergent Validity

Caitlin A. Stamatis^{1,2} · Jordana Muroff³ · Elizabeth S. Bocanegra⁴ · Carolyn I. Rodriguez^{5,6} · Kiara R. Timpano¹

Accepted: 10 May 2021 / Published online: 29 June 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Hoarding disorder (HD) involves extreme difficulties discarding possessions and significant clutter in living areas. Although hoarding occurs worldwide, cross-cultural research remains in nascent stages, hampered in part by a lack of validated measures in non-English languages. We aimed to validate a Spanish translation of the Hoarding Rating Scale (HRS), a widely used measure of core HD symptoms. Our sample ($N = 736$) included participants responding in English ($n = 548$; 45.4% female; 7.9% Latinx) or Spanish ($n = 188$; 46.3% female; 79.9% Latinx) to questionnaires via Amazon's Mechanical Turk. An item response theory (IRT) approach was used to test differential item functioning (DIF) of the English and Spanish HRS. We also examined convergent validity of each language version with other HD. Initial comparisons revealed that hoarding symptoms were elevated in the Spanish-speaking sample compared to the English-speaking sample. DIF tests flagged the clutter item for potential bias (McFadden's $\beta = .069$), but closer examination revealed that the impact was negligible. The Spanish HRS was significantly linked with other hoarding measures (Saving Inventory-Revised: $\beta = .497$, $p < .001$; Obsessive-Compulsive Inventory-Revised-Hoarding Subscale: $\beta = .329$, $p = .008$), controlling for mood, anxiety, stress, and non-hoarding OCD symptoms. However, the Spanish HRS was not significantly associated with Clutter Image Rating scores. Findings supported the utility of the HRS to measure of HD symptoms in Spanish speakers, though cross-linguistic assessment of clutter and the applicability of clinical cutoffs with Spanish-speaking samples merits further study.

Keywords Hoarding · Hoarding Rating Scale · Differential item functioning · Spanish translation · Item-response theory · Assessment · Validation

Introduction

Hoarding disorder (HD) constitutes a debilitating psychiatric condition and significant public health concern worldwide (Saxena et al., 2011; Tolin et al., 2008a, b). The core features of HD include extreme difficulty discarding everyday possessions and significant clutter in living areas. The majority of patients also endorse excessive acquiring (American Psychiatric Association, 2013). Hoarding symptoms are dimensionally distributed (Timpano et al., 2013), ranging from normative collecting tendencies to debilitating hoarding behavior, which occurs in approximately 2.5 percent of the population (Postlethwaite et al., 2019). At clinical levels, hoarding causes substantial impairment in personal, social, and financial functioning for the affected individual, with

cluttered living conditions also posing health risks at the family and community levels (Tolin et al., 2008a, b).

Although hoarding occurs worldwide (American Psychiatric Association, 2013; Fontenelle & Grant, 2014), cross-cultural research remains in nascent stages (Fernández de la Cruz et al., 2016). With several notable exceptions (Fernández de la Cruz et al., 2016; Nordsletten et al., 2018; Timpano et al., 2015; Tsuchiyagaito et al., 2017), the majority of phenomenological studies of HD have relied on predominantly Caucasian samples from the U.S. and/or Europe. This dearth of research across cultures poses the risk of a myopic view of hoarding, limiting awareness of key factors that may contribute to differences in vulnerability and treatment response (Fernández de la Cruz et al., 2016). Within the U.S., there is a particular need to better understand hoarding among Hispanic/Latinx individuals, who represent the second-largest racial or ethnic group after white non-Hispanics and the second fastest-growing group after Asian Americans (Noe-Bustamante et al., 2020).

✉ Kiara R. Timpano
kiaratimpano@gmail.com

Extended author information available on the last of the article

A necessity in advancing cross-cultural research is the availability of validated measures in languages other than English (Geisinger, 1994). A common self-report measure of hoarding, the Saving Inventory-Revised (SI-R; Frost et al., 2004), has been translated into several languages. To assess the adequacy of these translations, validation studies have focused on demonstrating divergent and convergent validity, respectively, by examining associations the translated SI-R with validated measures of mood and anxiety symptoms, as well as other hoarding measures. The self-report questionnaires that are most often used for the latter purposes are those that assess HD symptoms in the context of obsessive–compulsive disorder (OCD), such as the Obsessive–Compulsive Inventory-Revised (OCIR; Foa et al., 2002). Using these methods, the SI-R has been successfully adapted to Chinese (Timpano et al., 2015), Italian (Melli et al., 2013), German (Mueller et al., 2009), Portuguese (Fontenelle et al., 2010), and Spanish (Tortella-Feliu et al., 2006).

At the same time, research on linguistic adaptations of other measures of hoarding remains limited, particularly for the Hoarding Rating Scale (HRS; Tolin et al., 2010). The HRS is a widely used, 5-item measure of hoarding with several advantages for cross-cultural research. Namely, the HRS is brief, can be administered in self-report or interview form, maps directly onto the diagnostic criteria for HD, and boasts high reliability. The single extant study that formally adapted the HRS cross-culturally established a Japanese translation, which exhibited strong psychometric properties (Tsuchiyagaito et al., 2017). Although no studies have validated a Spanish translation of the HRS, ad hoc translations of the HRS are already being used in clinical and research settings to assess hoarding in Spanish-speakers (Nordsletten et al., 2018; Rodriguez et al., 2012). Though limited by a small sample ($n < 25$ per group), Nordsletten and colleagues (2018) found evidence that HRS scores did not significantly differ across Spanish and English-speaking samples. The fact that the Spanish HRS is already being used for community and cross-cultural research—despite a lack of evidence regarding its psychometric properties in Spanish-speaking samples—underscores the urgent need for a more formal assessment of its reliability and validity. If researchers hope to use the Spanish HRS to measure potential cross-cultural differences, there is a critical need to first establish that the Spanish language HRS represents as valid a measure of the hoarding construct as its English counterpart.

In addressing the need for a validated Spanish HRS, it is important to consider methodological choices, as well as the limitations of prior translation studies of related measures. Translating and validating a measure from English to another language is complex. To execute a successful translation, a native (bilingual) speaker will translate the measure into the focal language, then a second person will

back-translate the translation into the original language (Geisinger, 1994). Discrepancies can be resolved by an editorial board, ideally comprised of bilingual individuals from a range of backgrounds. At this point, researchers assess the validity of the translated measure. Commonly, researchers deem a translated measure “valid” if it is strongly associated with previously validated measures of the same construct (convergent validity) and less correlated with measures of distinct but related constructs (divergent validity). This methodology, which fueled the adaptations of the SI-R and OCIR (Fontenelle et al., 2010; Fullana et al., 2005; Melli et al., 2013; Timpano et al., 2015; Tortella-Feliu et al., 2006), represents an important but insufficient step. Prior to assessing convergent and divergent validity, there is a need to understand potential bias at the item level, which indicates that some factor unrelated to the construct of interest (e.g., hoarding) is causing one group to over- or under-report symptoms (Hui & Triandis, 1985). If items are biased—meaning they are not measuring a construct the same way across languages—tests of broader validity become uninterpretable (Thissen et al., 1993). Only when item-level equivalence is established can a researcher turn to examine other types of validity (e.g., convergent validity) and begin to test whether there are true group differences.

Item response theory (IRT) models represent an optimal approach to addressing questions about psychometric properties and potential item-level bias in the HRS. IRT posits that the response to an item (e.g., of the HRS) is attributable to both person-related characteristics (e.g., the level of hoarding symptoms) and item-related characteristics (e.g., how well the item differentiates between persons with high and low levels of hoarding). From an IRT perspective, the underlying levels of the latent trait (θ) should correspond with the likelihood of endorsing an item measuring that trait. For example, individuals with high levels of hoarding should be more likely to endorse higher categories of each HRS item. Particularly relevant to our study, IRT methods permit the measurement of differential item functioning (DIF). DIF occurs when persons who come from different groups—but who have the same level of θ (and should thus have the same likelihood of endorsing an item)—have different probabilities of item endorsement (Thissen et al., 1993). DIF tests capture potential bias in each individual item of a measure, thus representing an ideal choice for evaluating the HRS translation and original version.

In the present study, we took a stepwise approach to validating a Spanish translation of the HRS. Along with translation and back-translation techniques (Sousa & Rojjanasirrat, 2011), we compared our translation to one of the extant ad hoc Spanish translations (Rodriguez et al., 2012) and reconciled any differences to arrive at the final Spanish HRS. We examined basic psychometric properties of the HRS across each language version (i.e., response category functioning,

reliability, and dimensionality) in a community sample, followed by tests of DIF. Given prior findings of measurement invariance of other hoarding-specific scales across languages (Tortella-Feliu et al., 2006), we anticipated that the HRS would function similarly across groups and that tests of DIF would be nonsignificant. Our second objective was to test convergent and divergent validity. We first assessed zero-order correlations of the HRS with other measures of HD and related symptoms (anxiety, depression, and non-hoarding OCD). Next, in regression models, we considered the strength of the relationship of each language version of the HRS with other validated hoarding measures, while controlling for comorbid symptoms (depression, anxiety, stress, and OCD). By indicating whether the HRS is more strongly predicted by a given measure of hoarding (e.g., the SI-R) than by measures of common comorbidities in the same model, these analyses provided specificity. We predicted that both the English and Spanish versions of the HRS would be significantly associated with the other hoarding measures beyond the effects of common comorbidities.

Methods

Participants and Procedures

We recruited a sample of English- and Spanish-speaking participants ($N = 767$; English $n = 554$; Spanish $n = 213$) through Amazon's Mechanical Turk (MTurk) system. All participants provided informed consent and received monetary compensation. Prior to analysis, data were screened using established procedures for MTurk studies (e.g., Behrend et al., 2011) to ensure valid participant responses. In addition to screening the time taken to complete the survey, we examined responses to five randomly placed attention checks, in line with previous studies (e.g., Arditte et al., 2016). We excluded 31 participants for failing to pass the majority of our attention checks or completing the survey in less than 60% of the projected time, resulting in a final sample of 736 individuals.

Participants reported demographic information and completed a battery of questionnaires in English or Spanish based on language fluency. Only participants reporting that they were fluent in Spanish and comfortable completing a series of questionnaires in Spanish were administered the Spanish language version of the HRS. Primary English speakers ($n = 548$; 45.4% female; M age = 35.1 years, $SD = 11.1$ years; 7.9% Hispanic/Latinx) were sampled from across the U.S. Additional MTurk HITs were restricted to Texas, California, and Florida in order to bolster our Spanish-speaking sample (Spanish $n = 188$; 46.3% female; M age = 26.52 years, $SD = 8.94$ years; 79.9% Hispanic/Latinx).

Measures

Hoarding Rating Scale (HRS; Tolin et al., 2008a, b, 2010) The HRS is a five-item measure of hoarding symptoms that maps onto the core diagnostic criteria for HD, including difficulties discarding, clutter, excessive acquiring, distress, and impairment (Tolin et al., 2010; Tolin et al., 2008a, b). Items are rated on a 9-point Likert-type scale (0 = not at all; 2 = mildly; 4 = moderately; 6 = severely; 8 = extremely) according to severity over the prior week. The HRS can be administered as an interview or self-report, with both versions demonstrating high reliability (Tolin et al., 2010; Tolin et al., 2008a, b). As the present study was conducted online, we employed the self-report form. A score of 14 has been identified as the clinical cutoff (Tolin et al., 2010). Reliability of the HRS is reported for both the English and Spanish samples in the Results section below.

The HRS was translated into Spanish using a combined translation/back-translation and editorial board approach. This method is considered preferable to translation/back-translation alone, as the editorial board accounts for linguistic variation among different Spanish-speaking groups (Geisinger, 1994). First, a native Spanish speaker who was fully proficient in English translated the measure into Spanish. Next, the editorial board members (including individuals from Venezuela, Cuba, Mexico, Colombia, and Puerto Rico) reviewed the translation along with an existing ad hoc Spanish HRS and the English version. Feedback on discrepancies was incorporated, with an attempt to make the Spanish version easy to understand regardless of country of origin. Finally, a native English speaker proficient in Spanish back-translated the Spanish version into English. The back-translation and original English version were compared to ensure that the same information was being captured with each question. The Appendix contains the Spanish HRS, which has an estimated reading level of 7th–8th grade.

Saving Inventory-Revised (SI-R; Frost et al., 2004) The SI-R is a 23-item self-report measure of hoarding behaviors (Frost et al., 2004). Participants respond to items using a 5-point Likert scale (0 = none to 4 = almost all/extreme/very often). The SI-R includes three subscales that capture the core features of hoarding, which are difficulty discarding, clutter, and excessive acquiring. The SI-R exhibits strong psychometric properties in both clinical and nonclinical samples, and in both English (Coles et al., 2003; Frost et al., 2004) and Spanish (Tortella-Feliu et al., 2006). A clinical cutoff of 41 has been identified (Frost et al., 2004; Tolin et al., 2011). In our sample, there was good reliability for both the English ($\alpha = .94$) and Spanish ($\alpha = .96$) versions.

Clutter Image Rating (CIR; Frost et al., 2008) The CIR is a picture-based measure of clutter (Frost et al., 2008). It consists of three items (living room, bedroom, and kitchen), each of which contain nine images with progressively increasing levels of clutter. Participants are instructed to select the image for each room that best approximates the amount of clutter in that room of their home. Images are rated from 1 (least cluttered) to 9 (most cluttered), with a 4 representing clinically significant clutter. A total score is computed as the average of the scores from the three items. Previous studies have indicated that participant CIR ratings correspond well to clinician ratings of clutter (Frost et al., 2008; Tolin et al., 2010). Since the CIR is primarily a pictorial measure, we simply translated the instructions into Spanish using translation/back-translation approaches.

Depression, Anxiety, and Stress Scale (DASS; Crawford & Henry, 2003) The DASS is a 21-item self-report measure of depression (DASS-D), anxiety (DASS-A), and stress (DASS-S). Items are rated using a 4-point Likert scale ranging from 0 (did not apply to me at all) to 3 (applied to me very much or most of the time). The DASS displays strong psychometric properties in both English and Spanish (Bados et al., 2005; Crawford & Henry, 2003; Daza et al., 2002). A score of 14–20 indicates moderate depression, a score of 10–14 indicates moderate anxiety, and a score of 19–25 indicates moderate stress (Lovibond & Lovibond, 1996). In our sample, reliability was strong across subscales for the English (DASS-D: $\alpha = .94$; DASS-A: $\alpha = .87$; DASS-S: $\alpha = .91$) and Spanish (DASS-D: $\alpha = .93$; DASS-A: $\alpha = .90$; DASS-S: $\alpha = .90$) versions.

Obsessive–Compulsive Inventory-Revised (OCIR; Foa et al., 2002) The OCIR is an 18-item measure of OCD symptoms (Foa et al., 2002). Using a 5-point scale, participants rate the degree to which they have been bothered by each symptom over the past month. The OCIR contains six subscales: washing/contamination, checking, obsessions, neutralizing, ordering, and hoarding. Reliability of the OCIR has been established for both clinical (Abramowitz & Deacon, 2006; Foa et al., 2002) and nonclinical (Fullana et al., 2005) samples, in both English and Spanish (González et al., 2011). The clinical cutoff for the OCIR is 21 (Abramowitz & Deacon, 2006). For our study, we separated out the OCIR subscales into OCIR-hoarding (OCIR-H) and OCIR-non-hoarding (OCIR-NH), the latter of which was a composite of the other five subscales. In our sample, there was good reliability for both the English ($\alpha = .91$) and Spanish ($\alpha = .95$) versions of the OCIR.

Statistical Analyses

Model scripts are publicly available on RPubS.

Model Assumptions R and RStudio were used to assess psychometric properties of the HRS across both language groups. We employed the *itemanalysis* package (unpublished) for polytomous data to test response category functioning across the nine response options. We focused on the proportion of respondents selecting each category for each item, considering categories selected by less than 2 percent of the sample to be problematic (Cappelleri et al., 2014). The alpha function in the *psych* package was used to assess the reliability of the measure in each language (Revelle & Revelle, 2015). In addition, given the unidimensionality assumption of DIF tests, we tested dimensionality of both language versions of the HRS using modified parallel analysis. Specifically, the empirical second eigenvalue was compared to the 95th percentile of the second eigenvalue generated from 1000 simulated datasets with the same dimensions.

Differential Item Functioning (DIF) Prior to testing DIF, we compared model fit statistics for various polytomous IRT models, which indicated that the Graded Response Model (GRM) best fit the data (see Supplementary Materials Table S1 and Table S2 for details). All DIF analyses were run using R's *lordif* package (Choi et al., 2011), which employs the GRM, with English specified as the reference group. While several criteria can be used for DIF detection, we focused on McFadden's β , with a change of at least 5% in McFadden's β indicating slight to moderate DIF (Choi et al., 2011; Crane et al., 2004). We chose not to focus on the X^2 likelihood ratio test, as this metric has been demonstrated to overestimate DIF, nor on pseudo R^2 , which may underestimate meaningful DIF (Choi et al., 2011; Crane et al., 2007). To follow up on items showing DIF, we visually inspected various item trace plots to better characterize DIF as uniform or non-uniform (Walker, 2011). Uniform DIF occurs when respondents in one group systematically over- or under-endorse an item across different response categories and levels of the latent trait (θ ; i.e., level of hoarding). Non-uniform DIF occurs when bias is dependent upon levels of the latent trait (e.g., if respondents in one group over-endorse an item with respect to their level of the latent trait, but this only occurs at higher levels of severity). Item trace plots map the likelihood of endorsing a given response category across estimated levels of the latent trait. If the trace lines are disordinal or crossing, there may be non-uniform DIF.

Convergent and Divergent Validity We first examined correlations between the HRS total and other measures of hoarding symptoms (SI-R, CIR, and OCIR-H), as well as measures of depression (DASS-D), anxiety (DASS-A), stress (DASS-S), and non-hoarding OCD symptoms (OCIR-NH), with separate correlation matrices computed for each language group.

We followed these analyses with a series of regressions to examine the specificity of relationships, in line with previous studies (Tortella-Feliu et al., 2006). Separate models were examined for the English and Spanish subsamples. In these models, the predictors included one of the other hoarding measures (i.e., SI-R, CIR, or OCIR-H) and all comorbid symptoms (DASS-D, DASS-A, DASS-S, and OCIR-NH), with the HRS as the outcome. This type of analysis allowed for the examination of whether hoarding symptoms as assessed by the HRS were more strongly associated with other measures of hoarding (i.e., convergent validity), compared to comorbid symptoms (i.e., divergent validity), controlling for covariance between measures. For the SI-R analyses, we used a similar framework to look not only at the total score but also at the individual subscale scores (SI-R-discarding, SI-R-acquiring, and SI-R-clutter) in relationship to individual HRS items (HRS-discarding, HRS-acquiring, and HRS-clutter). For example, we tested a model where the outcome was the HRS-discarding item, and the predictors were the SI-R-discarding subscale score and covariates (DASS-D, DASS-A, DASS-S, and OCIR-NH). Finally, for the CIR, we looked at relationships with both the HRS total

score and also the individual HRS clutter item, again controlling for the same four covariates in each model.

Results

Basic sample descriptors, along with means and standard deviations for study measures, are presented in Table 1. Both the Spanish and English samples demonstrated relatively similar ranges across questionnaires. Independent samples t-tests were used to compare the means for each of the measures shown in Table 1. The Spanish sample was significantly younger than the English sample and scored significantly higher on all HRS items, HRS total score, SI-R clutter, SI-R acquiring, CIR, and OCIR-NH. The English sample scored significantly higher on DASS-D.

We next examined the percentage of each sample meeting the clinical cutoff on each measure. The percentage scoring in the moderate range or higher for the DASS depression and stress scales was comparable across the two samples. For the depression subscale, 32.97% of the English sample and 30.10% of the Spanish sample scored in the moderate range

Table 1 Comparison of item-level means and standard deviations (SD) for demographic and clinical characteristics, including the adjusted response-scale English and Spanish versions of the HRS

Measure	Spanish Sample	Spanish Range	English Sample	English Range	Statistical Comparison
Demographics					
Age	26.52 (8.94)	18.0–72.0	35.09 (11.08)	19.0–73.4	$t = -11.38, p < .001$
% Female	46.3%	–	45.4%	–	$\chi^2 = .00, p = .963$
HRS Items					
	M (SD)		M (SD)		
HRS 1 – Clutter	2.57 (2.01)	1–7	2.03 (1.51)	1–7	$t = 3.35, p < .001$
HRS 2 – Difficulties Discarding	2.98 (2.04)	1–7	2.37 (1.63)	1–7	$t = 3.67, p < .001$
HRS 3 – Acquiring	2.99 (2.00)	1–7	2.14 (1.54)	1–7	$t = 5.31, p < .001$
HRS 4 – Distress	3.07 (1.98)	1–7	2.01 (1.49)	1–7	$t = 6.74, p < .001$
HRS 5—Impairment	2.80 (2.10)	1–7	1.79 (1.38)	1–7	$t = 6.11, p < .001$
HRS Total Score	14.41 (8.98)	5–35	10.35 (6.56)	5–35	$t = 5.70, p < .001$
Other Hoarding Measures					
SI-R – Clutter	6.73 (7.51)	0–23	4.39 (5.24)	0–20	$t = 2.46, p = .015$
SI-R – Difficulties Discarding	6.92 (5.61)	0–18	5.84 (5.18)	0–20	$t = 1.36, p = .174$
SI-R – Acquiring	7.27 (5.83)	0–21	4.75 (4.35)	0–23	$t = 3.34, p = .001$
CIR	2.34 (1.60)	1.0–6.3	1.66 (.72)	1.0–5.3	$t = 3.81, p < .001$
OCIR-H	2.65 (2.87)	0–12	2.00 (2.18)	0–8	$t = 1.89, p = .059$
Non-Hoarding Measures					
DASS-D	7.51 (9.44)	0–32	9.85 (10.98)	0–42	$t = -2.16, p = .032$
DASS-A	7.59 (8.83)	0–30	6.22 (7.81)	0–40	$t = 1.41, p = .162$
DASS-S	9.48 (9.23)	0–32	10.46 (9.50)	0–42	$t = -.94, p = .351$
OCIR-NH	12.84 (12.17)	0–48	7.44 (8.42)	0–38	$t = 3.86, p < .001$

HRS Hoarding Rating Scale (adjusted response scale), SI-R Saving Inventory-Revised, CIR Clutter Image Rating, OCIR-H Obsessive–Compulsive Inventory-Revised – Hoarding subscale, OCIR-NH Obsessive–Compulsive Inventory-Revised – Non-Hoarding subscale, DASS-D Depression Anxiety Stress Scale – Depression subscale, DASS-A Depression Anxiety Stress Scale – Anxiety subscale, DASS-S Depression Anxiety Stress Scale – Stress subscale

* $p < .05$; ** $p < .01$; *** $p < .001$

or higher. For the stress subscale, 18.50% of the English sample and 18.28% of the Spanish sample scored in the moderate range or higher. In contrast, a greater proportion of the Spanish sample scored above the clinical cutoff on DASS anxiety and OCI-R symptoms. For the DASS anxiety subscale, 26.74% of the English sample and 35.48% of the Spanish sample scored in the moderate range or higher. For the OCIR, 15.63% of the English sample and 31.97% of the Spanish sample scored above the clinical cutoff. This pattern was mirrored for all of the hoarding measures, but with an even larger discrepancy between samples. On the HRS, 11.31% of the English sample scored above or equal to the clinical cutoff of 14, compared to 35.10% of the Spanish sample. On the SIR, 5.21% of the English sample scored above or equal to the clinical cutoff of 41, compared to 23.08% of the Spanish sample.

Basic Psychometric Features

Response Category Functioning Results from tests of the HRS response category functioning indicated issues with

response category selection proportion across both language versions. Specifically, an examination of the proportion selected for each of the nine original response categories demonstrated issues with category selection at the higher end, where for all items, less than 2 percent of respondents chose categories 7 or 8 (Fig. 1a). Consequently, response categories 6, 7, and 8 were collapsed into one category, resulting in a revised 7-response-options version of the HRS. When the response category selection proportions were recomputed for each item using the collapsed data, all 7 response categories were selected by at least 2 percent of subjects (Fig. 1b).

Reliability and Dimensionality. Both the English and Spanish scales exhibited excellent reliability (English HRS: $\alpha = .92$; Spanish HRS: $\alpha = .93$) and met the assumption of unidimensionality, as evidenced by large first eigenvalues (English: 4.13; Spanish: 4.16), compared to second eigenvalues $\leq .35$. From modified parallel analysis, the 95th percentile of the sampling distribution of the second

Fig. 1 Response category selection for 9- and 7-option versions of the HRS. **(a)** Response category selection for the 9-response option version of the HRS. **(b)** Response category selection for the 7-response option version of the HRS. Black bar indicates the cutoff of a minimum of 2 percent of respondents choosing a given response category for a given item

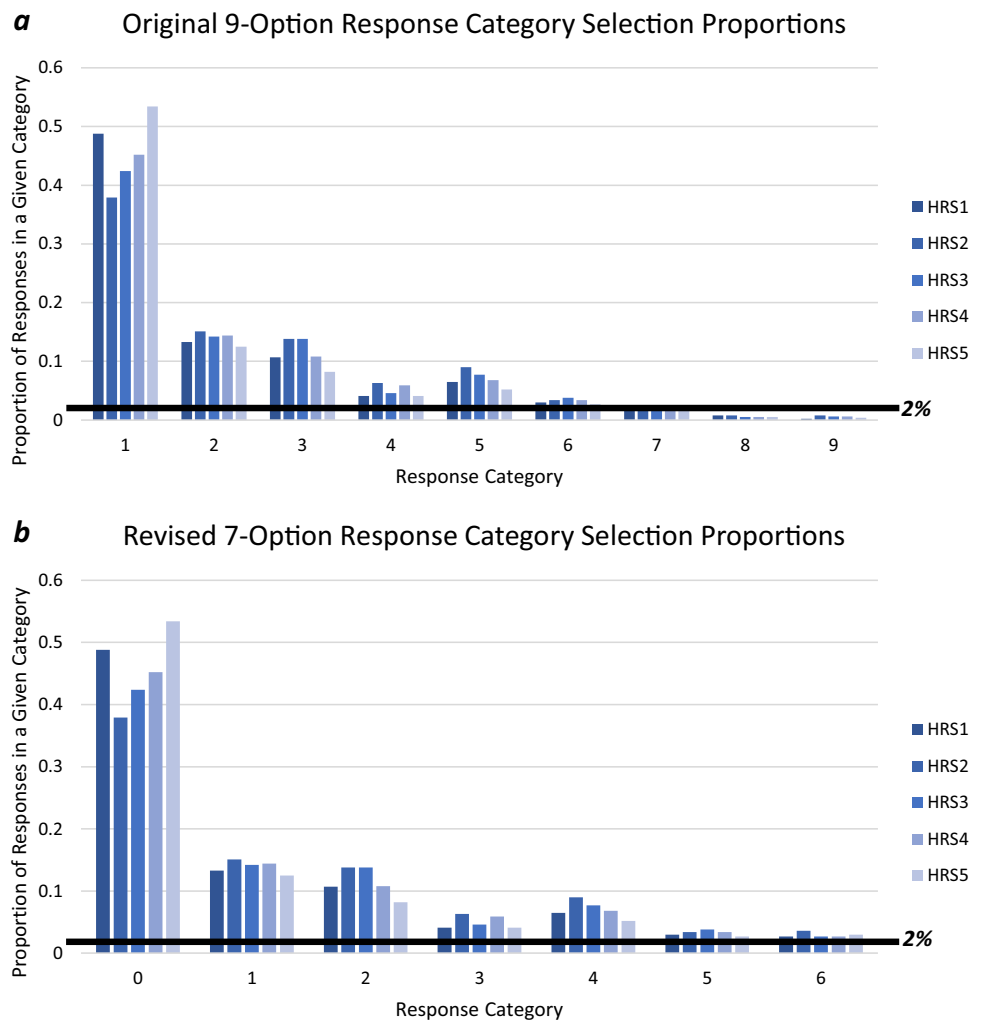
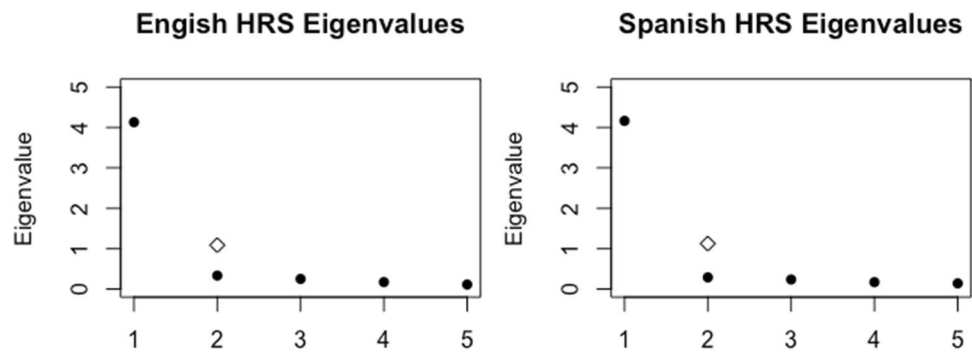


Fig. 2 Eigenvalues for English and Spanish versions of the HRS. ● indicates empirical eigenvalues; ◇ indicates 95th percentile of sampling distribution of 2nd eigenvalue from 1000 simulated datasets



eigenvalue was significantly higher than the empirical second eigenvalue, further supporting unidimensionality (Fig. 2).

DIF Tests

The first item of the HRS, which captures clutter, exhibited DIF across language groups, with a McFadden’s β of .0686. However, the impact of DIF on the clutter item and overall HRS score was negligible (Fig. 3), particularly when weighted by density (see “Item trace and information plots” section below for additional information). More detailed DIF results can be found in Supplementary Table S3. Latent trait distributions by language group can be found in Supplementary Fig. S1. The remaining four HRS items did not show evidence of DIF.

Item Parameter Estimates Model results were further examined to aid in interpreting DIF for the clutter item. We focused on the a parameter, which controls steepness of operating characteristic and response curves, as well as the δ step parameter (Table 2), both of which were estimated separately across languages for the DIF item. Differences in the a parameter (English $a = 4.183$; Spanish $a = 3.375$) indicated that the clutter item better discriminated between levels of hoarding for the English as compared to the Spanish group (Fig. 3). The Spanish version had consistently higher δ parameters for all response categories of item 1. This indicates that individuals with the same level of the latent trait in English and Spanish were marginally less likely to endorse clutter in the Spanish version.

Fig. 3 Plots of HRS item 1. Item 1, which measures clutter, exhibits uniform DIF. (a) ICCs for English and Spanish groups. (b) Absolute difference in ICCs for English and Spanish groups. (c) Item response functions for each response category by language group. (d) Difference in ICCs weighted by score distribution for focal group (Spanish), indicating very minimal impact just above average scores

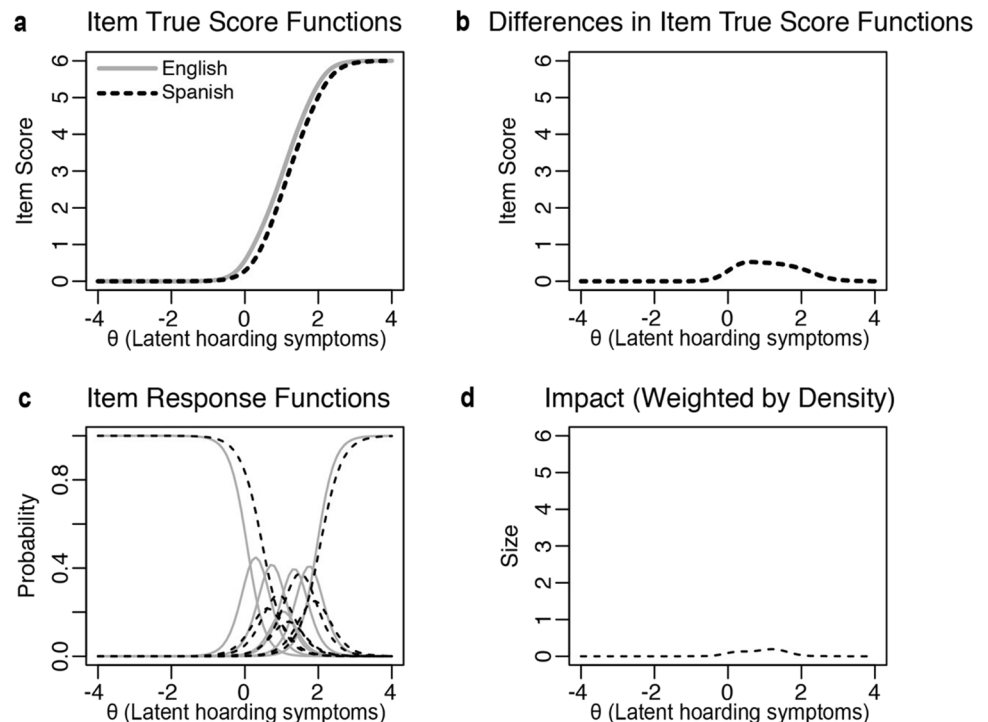


Table 2 Item parameter estimates for the English and Spanish HRS

HRS Item	a	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6
HRS1-E	4.183	0.062	0.524	0.948	1.145	1.547	1.962
HRS1-S	3.375	0.484	0.745	1.084	1.272	1.741	2.044
HRS2	2.832	-0.205	0.312	0.834	1.118	1.652	2.012
HRS3	3.260	-0.054	0.400	0.929	1.150	1.637	2.087
HRS4	4.422	0.044	0.501	0.913	1.181	1.587	1.944
HRS5	5.701	0.284	0.683	1.030	1.241	1.580	1.836

a =slope; δ =step parameters. For item 1 (DIF item), parameters were estimated separately by language group; “-E” indicates English, and “-S” indicates Spanish

Item Trace and Information Plots We next examined the clutter item trace lines by group (Fig. 3), which indicated evidence of primarily uniform DIF. The Spanish trace lines were slightly shifted to the right, indicating that Spanish speakers were less likely to endorse the higher response categories even after putting both groups on the same metric. There was minimal indication of non-uniform DIF for item 1, as the shifts in item difficulty were relatively consistent across the latent trait continuum, except for disordinal or crossing lines at the highest levels of the latent trait continuum (category 6). Notably, the impact of DIF was minimal (Fig. 3), particularly given that few individuals endorsed levels of the latent trait at the point where DIF was strongest. Test characteristic curves for all items and the clutter item, as well as plots of individual-level DIF impact, can be found in Supplementary Figs. S2 and S3.

Mean Scores After Calibration After linking groups on the same metric and recoding scores for each item to a 0–7 scale (Table 1), the average HRS score for Spanish speakers was significantly higher than that of the English group, $t(258.83) = 5.70, p < .001$; Spanish $M = 14.41$ ($SD = 8.98$); English $M = 10.35$ ($SD = 6.56$). Given that the two groups differed on mean age, we further examined whether HRS

score was impacted by age by regressing HRS total score on group and age. Controlling for age, the Spanish group had significantly higher HRS scores than the English group ($\beta = .49, 95\% \text{ CI: } -.260-.720, p < .001, \eta^2_{\text{partial}} = .027$). Age was not a significant predictor of HRS scores ($\beta = -.04, 95\% \text{ CI: } -.120-.040, p = .314, \eta^2_{\text{partial}} = .002$), and it therefore was not included as a covariate in subsequent models.

Convergent and Divergent Validity

Using the modified response-scale versions of the HRS, we first examined correlations with all other symptom measures. In both English and Spanish, the HRS generally exhibited stronger relationships with other measures of hoarding compared to measures of comorbid constructs, including depression, anxiety, stress, and non-hoarding OCD symptoms (Table 3).

We next conducted a series of regressions to more closely examine convergent validity with the additional hoarding symptom measures, controlling for any covariance with the commonly comorbid symptom measures. The specific regression models included the following: 1) HRS total regressed on SI-R total, DASS-D, DASS-A, DASS-S, OCIR-NH; 2) HRS clutter item regressed on SI-R clutter subscale,

Table 3 Zero-order correlations between symptom measures across English (grey-shaded areas) and Spanish subsamples

	1. HRS	2. SI-R	3. CIR	4. OCIR-H	5. OCIR-NH	6. DASS-D	7. DASS-A	8. DASS-S
1. HRS	-	.80***	.70***	.75***	.47***	.40***	.48***	.46***
2. SI-R	.82***	-	.68***	.84***	.54***	.41***	.38***	.58***
3. CIR	.44***	.53***	-	.66***	.34***	.33***	.30**	.41***
4. OCIR-H	.60***	.77***	.48***	-	.47***	.39***	.30**	.53***
5. OCIR-NH	.65***	.79***	.55***	.84***	-	.50***	.53***	.64***
6. DASS-D	.68***	.74***	.51***	.64***	.73***	-	.64***	.77***
7. DASS-A	.73***	.77***	.48***	.67***	.79***	.84***	-	.77***
8. DASS-S	.71***	.71***	.48***	.64***	.70***	.83***	.83***	-

HRS Hoarding Rating Scale, SI-R Saving Inventory-Revised, CIR Clutter Image Rating, OCIR-H Obsessive-Compulsive Inventory-Revised – Hoarding subscale, OCIR-NH Obsessive-Compulsive Inventory-Revised – Non-Hoarding subscale, DASS-D Depression Anxiety Stress Scale – Depression subscale, DASS-A Depression Anxiety Stress Scale – Anxiety subscale, DASS-S Depression Anxiety Stress Scale – Stress subscale

*** $p < .001$

DASS-D, DASS-A, DASS-S, OCIR-NH; 3) HRS discarding item regressed on SI-R discarding subscale, DASS-D, DASS-A, DASS-S, OCIR-NH; 4) HRS acquiring item regressed on SI-R acquiring subscale, DASS-D, DASS-A, DASS-S, OCIR-NH; 5) HRS total regressed on CIR total, DASS-D, DASS-A, DASS-S, OCIR-NH; 6) HRS clutter item regressed on CIR total, DASS-D, DASS-A, DASS-S, OCIR-NH; 7) HRS total regressed on OCIR hoarding subscale, DASS-D, DASS-A, DASS-S, OCIR-NH. Each model was tested separately for the English and Spanish groups.

HRS Convergent Validity With the SI-R Controlling for the DASS subscales and OCIR-NH, we found that the HRS was significantly associated with hoarding on the SI-R for both the English ($\beta = .756$, 95% CI: .599-.914, $p < .001$, $\eta^2_{\text{partial}} = .504$) and Spanish ($\beta = .497$, 95% CI: .295-.699, $p < .001$, $\eta^2_{\text{partial}} = .220$) groups, with the English effect size more than twice the magnitude of the Spanish effect size. In terms of associations with covariates, the HRS was related to OCIR-NH in the Spanish group ($\beta = .228$, 95% CI: .022-.434, $p = .030$, $\eta^2_{\text{partial}} = .054$), but not the English group ($\beta = .000$, 95% CI: -.164-.165, $p = .998$, $\eta^2_{\text{partial}} = .000$).

Next, we conducted three follow-up regressions, with each individual HRS item (clutter, difficulty discarding, and excessive acquiring) serving as the respective DV and the respective SI-R subscales (clutter, difficulty discarding, and acquiring) replacing the SI-R total score as the primary predictor of interest. The HRS clutter item was significantly associated with the SI-R clutter subscale across both the English ($\beta = .655$, 95% CI: .477-.834, $p < .001$, $\eta^2_{\text{partial}} = .372$) and Spanish ($\beta = .367$, 95% CI: .098-.635, $p = .008$, $\eta^2_{\text{partial}} = .080$) groups. The English effect size was over four times the magnitude of the Spanish effect size. The HRS clutter item was differentially related to covariates across the two language groups. Namely, the HRS clutter item was significantly related to DASS-A for the English ($\beta = .234$, 95% CI: .015-.452, $p = .036$, $\eta^2_{\text{partial}} = .048$) but not Spanish ($\beta = .206$, 95% CI: -.132-.544, $p = .228$, $\eta^2_{\text{partial}} = .017$) group.

The HRS difficulty discarding item was significantly associated with the SI-R difficulty discarding subscale for the English ($\beta = .793$, 95% CI: .623-.964, $p < .001$, $\eta^2_{\text{partial}} = .487$) and Spanish ($\beta = .328$, 95% CI: .116-.540, $p = .003$, $\eta^2_{\text{partial}} = .100$) groups. The effect size was substantially larger for the English compared to the Spanish group. Controlling for SI-R difficulty discarding, the HRS discarding item was also significantly related to DASS-S in the Spanish group ($\beta = .366$, 95% CI: .080-.652, $p = .013$, $\eta^2_{\text{partial}} = .071$), but not in the English group ($\beta = .019$, 95% CI: -.261-.300, $p = .892$, $\eta^2_{\text{partial}} = .000$). Notably, the magnitude of the association between HRS-discarding and DASS-S was comparable to that of HRS-discarding and

SI-R-discarding in the Spanish group, challenging divergent validity hypotheses.

The HRS acquiring item was significantly associated with the SI-R acquisition subscale for the English ($\beta = .711$, 95% CI: .547-.875, $p < .001$, $\eta^2_{\text{partial}} = .452$) and Spanish ($\beta = .493$, 95% CI: .270-.716, $p < .001$, $\eta^2_{\text{partial}} = .186$) groups, with more than twice as large of an effect for the English group.

HRS Convergent Validity With the CIR Relationships with clutter on the CIR were less consistent. We first examined the model with HRS total as the dependent variable and CIR as the predictor, controlling for DASS and OCIR-NH. We found that while there was a significant association between CIR and HRS total for the English group ($\beta = .546$, 95% CI: .392-.700, $p < .001$, $\eta^2_{\text{partial}} = .356$), this was not the case for the Spanish group ($\beta = -.005$, 95% CI: -.163-.153, $p = .950$, $\eta^2_{\text{partial}} = .000$). In terms of covariates in the CIR model, OCIR-NH was significantly linked with HRS total in the Spanish group ($\beta = .430$, 95% CI: .206-.655, $p < .001$, $\eta^2_{\text{partial}} = .145$), but not the English group ($\beta = .167$, 95% CI: -.012-.347, $p = .067$, $\eta^2_{\text{partial}} = .037$).

A similar pattern of findings emerged for a model with the HRS clutter item as the DV. The HRS clutter item was significantly linked with CIR scores for the English group ($\beta = .667$, 95% CI: .508-.826, $p < .001$, $\eta^2_{\text{partial}} = .437$), but not for the Spanish group ($\beta = .077$, 95% CI: -.109-.264, $p = .410$, $\eta^2_{\text{partial}} = .008$). With regard to covariates, the HRS clutter item was significantly associated with the OCIR-NH subscale for the Spanish group ($\beta = .366$, 95% CI: .102-.631, $p = .007$, $\eta^2_{\text{partial}} = .081$), but not for the English group ($\beta = -.080$, 95% CI: -.265-.105, $p = .392$, $\eta^2_{\text{partial}} = .008$).

HRS Convergent Validity With OCIR-hoarding HRS scores were significantly related to the OCIR hoarding subscale for both the English ($\beta = .662$, 95% CI: .502-.822, $p < .001$, $\eta^2_{\text{partial}} = .429$) and Spanish ($\beta = .329$, 95% CI: .089-.569, $p = .008$, $\eta^2_{\text{partial}} = .080$) groups, with the English effect size more than five times the magnitude of the Spanish effect size. For both language groups, none of the covariates (DASS-D, DASS-A, DASS-S, and OCIR-NH) were significantly associated with HRS scores.

Discussion

The present study is the first to examine the psychometric performance of the Spanish HRS as a measure of hoarding symptoms. We employed an IRT approach to investigate differential item functioning of the HRS items across English and Spanish language versions, then assessed convergent validity of the Spanish and English HRS with validated measures of hoarding, controlling for common

comorbidities. Overall, results supported the Spanish HRS as a valid measure of hoarding symptoms. Psychometric evidence indicated similar performance of the Spanish HRS to the English HRS at the basic item level. While DIF tests flagged evidence of potential bias in the clutter item, the bias was negligible in terms of the magnitude of its effect on the clutter item and overall HRS score. Moreover, the Spanish translation of the HRS exhibited high convergent validity across most indices, although the magnitudes of these associations were weaker than those of the English HRS.

The Spanish HRS conformed to model assumptions, performing similarly to the English version across basic psychometric tests. Reliability was high across both versions, with a Cronbach's α of .92 for the English HRS and .93 for the Spanish HRS, suggesting that the items capture significant variance in the latent hoarding construct regardless of response language. Our findings also supported a unidimensional structure of the HRS items for both languages, in line with prior psychometric research on the English scale (Tolin et al., 2010; Tolin et al., 2008a, b). Taken together, these results suggest that the HRS is similarly reliable across languages, and the HRS can be treated as measuring a single underlying construct.

In line with study hypotheses, the Spanish HRS was associated with almost all other measures of hoarding, even after controlling for depression, anxiety, stress, and non-hoarding OCD symptoms. This pattern of results suggests that the Spanish HRS exhibits convergent validity, as does the English HRS (Tolin et al., 2010; Tolin et al., 2008a, b). Nevertheless, the magnitude of the associations between the Spanish HRS and related hoarding measures were weaker than those found for the English HRS. While weaker effect sizes may reflect greater variability in terms of how well the Spanish HRS captures hoarding, the smaller sample size for the Spanish group may also have contributed to the effect size discrepancy.

The Spanish version of the HRS also displayed weaker discriminant validity as compared to its English counterpart. The Spanish HRS total score was significantly associated with the OCIR non-hoarding subscale across various regression models, which was not the case for the English HRS. Moreover, the HRS discarding item was significantly associated with the DASS stress subscale for the Spanish group, and the magnitude of the association with DASS-S was comparable to that with the SI-R discarding subscale. Though unexpected, this finding aligns with evidence of weaker divergent validity for the Spanish version of the SI-R (Tortella-Feliu et al., 2006). The authors of the Spanish SI-R similarly reported significant associations of the Spanish SI-R—both the total and subscale scores—with affective and non-hoarding OCD symptoms (Tortella-Feliu et al., 2006), raising questions about discriminant validity.

The most complex finding emerged with respect to Spanish language assessments of clutter, which surfaced as potentially problematic in several different instances. First, contrary to hypotheses, the HRS clutter was the one item that exhibited DIF, with a trend suggesting that Spanish-speaking respondents with the same level of hoarding symptoms would be slightly less likely to endorse clutter. At the same time, the impact of this DIF appeared to be negligible even at the item level (Fig. 3), and there was little impact on the overall test characteristic curves (Fig. S2). While the presence of DIF in the clutter item is therefore unlikely to have impacted the validity of this item or the scale overall in measuring hoarding symptoms (Choi et al., 2011; Crane et al., 2007), these findings should be interpreted in light of the regression analysis with the CIR. This pictorial measure of clutter demonstrated a weak relationship with the HRS total score and clutter item specifically in the Spanish and not in English. The CIR findings stood in contrast to other analyses of clutter-based measures, as the Spanish HRS clutter item was significantly related to the SI-R clutter subscale. Our findings raise intriguing implications about the clutter criterion and assessment. While few studies have considered the CIR cross-culturally, one report indicated that Brazilian participants endorsed lower levels of clutter on the CIR compared to individuals in Spain, England, and Japan (Nordsletten et al., 2018). Conceptualizations of clutter may vary according to cultural (e.g., collectivism; familism; stigma) and environmental (i.e., living arrangements, space) factors, suggesting a need for future studies to consider the generalizability of CIR images. Effects of age and living situation—two factors known to influence the age of onset and course of hoarding (Ayers et al., 2010; Grisham et al., 2006)—may also influence manifestations of clutter across groups.

Despite remaining questions about the measurement of clutter, the Spanish HRS functioned similarly overall to the English HRS in terms of appropriately categorizing respondents based on their levels of the latent hoarding trait. The validity of the Spanish HRS across multiple indices is an important finding, as several prior studies in community settings have relied on Spanish translations of the HRS without information about its psychometric performance (Nordsletten et al., 2018; Rodriguez et al., 2012). Notably, we found that even after equating the language groups on the same scale, the Spanish mean HRS score was approximately four points higher than that of English speakers. This finding, which merits replication in clinical samples, raises the question of why hoarding symptoms might be reported as higher among certain groups than others. It will be important for future research to continue to assess whether hoarding occurs at similar levels across the globe, as well as whether language influences reports of hoarding symptomatology among bilingual individuals, as it has been found to do with

other psychiatric disorders (Brown & Weisman de Mamani, 2017; Guttfreund, 1990).

The results of the present study should be considered in light of limitations, pointing to avenues for future research. First, we employed a convenience sample of MTurk participants to assess hoarding symptoms measured dimensionally. Although MTurk samples endorse relatively high rates of psychiatric conditions (Arditte et al., 2016), it will be important to replicate these procedures in samples of patients with a confirmed hoarding diagnosis. Second, our sample was unbalanced in terms of the number in the English and Spanish groups; in particular, the smaller n in our focal group renders the DIF findings tentative. Third, and relatedly, the groups differed significantly in terms of age, with the Spanish sample representing a younger cohort. Although age was not significantly related to HRS scores in our sample, future studies would benefit from assessing a more diverse array of participants in terms of ages represented, as patients tend to present with clinically significant hoarding later into adulthood.

Additional questions for future studies involve response option functioning of the 0–8 scale currently used in the HRS and clinical cutoffs for hoarding measures across languages. The HRS 0–8 response scale was problematic in our community sample, with fewer than 2% of respondents selecting categories 7 and 8 for each of the HRS items in both language versions. Moreover, there was considerable overlap in the distributions of the response categories 3, 4, and 5 on the latent trait continuum for all items in both English and Spanish. This is notable given that the jump from 3 to 4 on the 9-point clinical severity rating scale represents the boundary between subclinical symptoms and “caseness” (American Psychiatric Association, 2013). Our findings related to the HRS scaling are also interesting to consider in light of the overall greater levels of hoarding symptoms endorsed by the Spanish-speaking respondents. While levels of depression and stress were comparable between the two samples, we found that a larger proportion of the Spanish-speaking respondents reported scores above the clinical cutoff for the HRS and all of the SI-R subscales. It should be noted that both HRS means were higher than those reported in other MTurk samples ($M = 6.47$; Arditte et al., 2016), but the Spanish sample mean was substantially higher. These results recall those of Timpano et al. (2015), who found markedly and significantly higher scores on the Saving Beliefs Questionnaire administered in Chinese ($M = 63.67$) compared to English ($M = 32.29$). If our results replicate in clinical and community samples, researchers may consider the sources of discrepancies on hoarding measures administered in English and Spanish, such as cultural factors and

biases in measurement. While the former could suggest a need to consider different cutoffs across groups, the latter would underscore the importance of comprehensive validation studies.

Results of the present study represent a stepping-stone for cross-cultural research on hoarding. Our findings underscore the utility of the HRS as a brief measure of hoarding symptoms in Spanish speakers, with the English and Spanish versions exhibiting comparable psychometric properties and reliability. However, results demonstrated some evidence of DIF in the measurement of clutter on the HRS, as well as lower convergent validity of the Spanish HRS with clutter on the CIR. While Spanish-speaking participants generally reported higher levels of hoarding symptoms compared to the English-speaking reference group, future studies employing larger samples of Spanish-speaking participants, including individuals in countries outside of the U.S., will aid in clarifying the generalizability of our findings. With continued study of the hoarding construct through a broader lens, the field may achieve a greater understanding of linguistic and cultural factors that influence differences in the phenomenology, risk, and treatment-seeking behaviors relevant to hoarding symptoms.

Appendix

Escala De La Acumulación (EDLA)

Las siguientes 5 preguntas son sobre su experiencia con la acumulación compulsiva de objetos. Por favor, responda las preguntas con respecto a su experiencia en la SEMANA PASADA.

1. Debido al desorden (desorganización y amontonamiento) de sus cosas o a la cantidad de cosas que tiene, ¿qué tan difícil es para usted usar las habitaciones en su casa?

0	1	2	3	4	5	6	7	8
Nada difícil	Levemente difícil		Moderadamente difícil		Muy difícil		Extremadamente difícil	

2. ¿Qué tan difícil se le hace botar (deshacerse, reciclar, vender, regalar) cosas comunes que otras personas botarían?

0	1	2	3	4	5	6	7	8
Nada difícil	Levemente difícil		Moderadamente difícil		Muy difícil		Extremadamente difícil	

3. ¿Actualmente, qué tanto problema tiene recolectando cosas gratuitas o comprando más cosas de las que necesita o puede usar o pagar? [Use la escala a continuación.]

0	1	2	3	4	5	6	7	8
Ningún problema	Problema leve		Problema moderado		Mucho problema		Problema extremo	

0 = ningún problema

2 = problema leve, en ocasiones (menos de una vez a la semana) adquiero o compro cosas no necesarias, o adquiero pocas cosas innecesarias

4 = problema moderado, regularmente (una o dos veces por semana) adquiero o compro cosas no necesarias, o adquiero algunas cosas innecesarias

6 = mucho problema, frecuentemente (varias veces por semana) adquiero o compro cosas no necesarias, o adquiero muchas cosas innecesarias

8 = problema extremo, muy frecuente (diariamente) adquiero o compro cosas no necesarias, o adquiero un gran número de cosas innecesarias

4. ¿Hasta qué punto tiene usted angustia debido al reguero (desorden y amontonamiento) de sus cosas, o a la dificultad de botar (desechar) cosas o al problema de comprar o adquirir cosas?

0	1	2	3	4	5	6	7	8
Nada	Angustia leve		Angustia moderada		Mucha angustia		Angustia extrema	

5. ¿Hasta qué punto se afecta su vida (rutina diaria, trabajo/escuela, actividades sociales, actividades familiares, dificultades económicas) debido al reguero (desorden y amontonamiento) de sus cosas, a la dificultad para botarlas (desecharlas) o al problema de comprar o adquirir cosas?

0	1	2	3	4	5	6	7	8
Nada	Levemente afectada		Moderadamente afectada		Muy afectada		Extremadamente afectada	

Findings support the utility of the EDLA (Escala de la acumulación) to measure hoarding disorder symptoms in Spanish-speaking samples; however, caution is warranted in applying clinical cutoffs derived from English-speaking samples. Additional research is recommended to confirm the appropriateness of the cutoff of ≥ 14 for Spanish-speaking populations

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10862-021-09894-z>.

Authors' Contributions As first and senior authors, respectively, CS and KT took the lead in conceptualizing and writing the manuscript. CS,

JM, EB, CR, and KT all played roles in the design and implementation of the research study. EB, CS, and KT were responsible for data collection, and CS conducted the primary statistical analyses in the manuscript. All authors played a role in the management and security of study data. All authors contributed to and have approved the final manuscript.

Funding The authors did not receive financial support for the conduct of the research and/or preparation of the article.

Availability of Data and Material Deidentified data are available upon request from the authors.

Code Availability Model scripts are available on RPubS.

Declarations

Experiment Participants All study procedures were approved by the University of Miami Institutional Review Board, and all participants provided informed consent.

Conflicts of Interest Dr. Carolyn Rodriguez has served as a consultant for Allergan, BlackThorn Therapeutics, Rugen Therapeutics, and Epiodyne, receives research grant support from Biohaven Inc., and a stipend from APA Publishing for her role as Deputy Editor at The American Journal of Psychiatry. Caitlin A. Stamatis, Jordana Muroff, Elizabeth S. Bocanegra, and Kiara R. Timpano declare that they have no financial or other relationships relevant to the subject of this manuscript.

References

- Abramowitz, J. S., & Deacon, B. J. (2006). Psychometric properties and construct validity of the Obsessive-Compulsive Inventory—Revised: Replication and extension with a clinical sample. *Journal of Anxiety Disorders*, 20(8), 1016–1035.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Publishing.
- Arditte, K. A., Çek, D., Shaw, A. M., & Timpano, K. R. (2016). The importance of assessing clinical phenomena in Mechanical Turk research. *Psychological Assessment*, 28(6), 684. <https://doi.org/10/f8x5v9>
- Ayers, C. R., Saxena, S., Golshan, S., & Wetherell, J. L. (2010). Age at onset and clinical features of late life compulsive hoarding. *International Journal of Geriatric Psychiatry: A Journal of the Psychiatry of Late Life and Allied Sciences*, 25(2), 142–149. <https://doi.org/10/dsgcdc>
- Bados, A., Solanas, A., & Andrés, R. (2005). Psychometric properties of the Spanish version of depression, anxiety and stress scales (DASS). *Psicothema*, 17(4), 679–683.
- Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, 43(3), 800. <https://doi.org/10/dncq9f>
- Brown, C. A., & Weisman de Mamani, A. (2017). A comparison of psychiatric symptom severity in individuals assessed in their mother tongue versus an acquired language: A two-sample study of individuals with schizophrenia and a normative population. *Professional Psychology: Research and Practice*, 48(1), 1.
- Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes

- measures. *Clinical Therapeutics*, 36(5), 648–662. <https://doi.org/10/652>
- Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of statistical software*, 39(8), 1. <https://doi.org/10/gf4tx2>
- Coles, M. E., Frost, R. O., Heimberg, R. G., & Steketee, G. (2003). Hoarding behaviors in a large college sample. *Behaviour Research and Therapy*, 41(2), 179–194.
- Crane, P. K., Belle, G. van, & Larson, E. B. (2004). Test bias in a cognitive test: differential item functioning in the CASI. *Statistics in Medicine*, 23(2), 241–256. <https://doi.org/10/bvmbwr>
- Crane, P. K., Gibbons, L. E., Ocepok-Welikson, K., Cook, K., Cella, D., Narasimhalu, K., et al. (2007). A comparison of three sets of criteria for determining the presence of differential item functioning using ordinal logistic regression. *Quality of Life Research*, 16(1), 69.
- Crawford, J. R., & Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology*, 42(2), 111–131.
- Daza, P., Novy, D. M., Stanley, M. A., & Averill, P. (2002). The depression anxiety stress scale-21: Spanish translation and validation with a Hispanic sample. *Journal of Psychopathology and Behavioral Assessment*, 24(3), 195–205.
- Fernández de la Cruz, L., Nordsletten, A. E., & Mataix-Cols, D. (2016). Ethnocultural aspects of hoarding disorder. *Current Psychiatry Reviews*, 12(2), 115–123.
- Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The Obsessive-Compulsive Inventory: development and validation of a short version. *Psychological Assessment*, 14(4), 485. <https://doi.org/10/dkmzv8>
- Fontenelle, I. S., Rangé, B. P., Prazeres, A. M., Borges, M. C., Versiani, M., & Fontenelle, L. F. (2010). The Brazilian Portuguese Version of the Saving Inventory–Revised: Internal Consistency, Test-Retest Reliability, and Validity of a Questionnaire to Assess Hoarding. *Psychological Reports*, 106(1), 279–296.
- Fontenelle, L. F., & Grant, J. E. (2014). Hoarding disorder: a new diagnostic category in ICD-11? *Brazilian Journal of Psychiatry*, 36, 28–39.
- Frost, R. O., Steketee, G., & Grisham, J. R. (2004). Measurement of compulsive hoarding: saving inventory-revised. *Behaviour Research and Therapy*, 42(10), 1163–1182.
- Frost, R. O., Steketee, G., Tolin, D. F., & Renaud, S. (2008). Development and validation of the clutter image rating. *Journal of Psychopathology and Behavioral Assessment*, 30(3), 193–203.
- Fullana, M. A., Tortella-Feliu, M., Caseras, X., Andión, Ó., Torrubia, R., & Mataix-Cols, D. (2005). Psychometric properties of the Spanish version of the Obsessive-Compulsive Inventory—Revised in a non-clinical sample. *Journal of Anxiety Disorders*, 19(8), 893–903.
- Geisinger, K. F. (1994). Cross-cultural normative assessment: translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, 6(4), 304.
- González, A. E. M., Piqueras, J. A., & Marzo, J. C. (2011). Validación del inventario de obsesiones y compulsiones revisado (OCIR) para su uso en población adolescente española. *Anales De Psicología/annals of Psychology*, 27(3), 763–773.
- Grisham, J. R., Frost, R. O., Steketee, G., Kim, H.-J., & Hood, S. (2006). Age of onset of compulsive hoarding. *Journal of Anxiety Disorders*, 20(5), 675–686. <https://doi.org/10/b3snnd>
- Guttfreund, D. G. (1990). Effects of language usage on the emotional experience of Spanish-English and English-Spanish bilinguals. *Journal of Consulting and Clinical Psychology*, 58(5), 604.
- Hui, C. H., & Triandis, H. C. (1985). Measurement in cross-cultural psychology: A review and comparison of strategies. *Journal of Cross-Cultural Psychology*, 16(2), 131–152.
- Lovibond, S. H., & Lovibond, P. F. (1996). *Manual for the depression anxiety stress scales*. Psychology Foundation of Australia.
- Melli, G., Chiorri, C., Smurra, R., & Frost, R. O. (2013). Psychometric properties of the paper-and-pencil and online versions of the Italian Saving Inventory—Revised in nonclinical samples. *International Journal of Cognitive Therapy*, 6(1), 40–56.
- Mueller, A., Crosby, R. D., Frost, R. O., Leidel, B., Bleich, S., Glaesmer, H., et al. (2009). Fragebogen zum zwanghaften Horten (FZH)—Validierung der deutschen Version des Saving Inventory-Revised. [German Compulsive Hoarding Inventory (FZH)—Evaluation of the German version of the Saving Inventory-Revised.]. *Verhaltenstherapie*, 19(4), 243–250. <https://doi.org/10/dfjjsj>
- Noe-Bustamante, L., Lopez, M. H., & Krogstad, J. M. (2020). US Hispanic population surpassed 60 million in 2019, but growth has slowed. *Fact Tank, News in the Numbers*. Pew Research Center.
- Nordsletten, A. E., Fernández de la Cruz, L., Aluco, E., Alonso, P., López-Solà, C., Menchón, J. M., et al. (2018). A transcultural study of hoarding disorder: Insights from the United Kingdom, Spain, Japan, and Brazil. *Transcultural Psychiatry*, 55(2), 261–285.
- Postlethwaite, A., Kellett, S., & Mataix-Cols, D. (2019). Prevalence of hoarding disorder: A systematic review and meta-analysis. *Journal of Affective Disorders*, 256, 309–316.
- Revelle, W., & Revelle, M. W. (2015). Package ‘psych.’ *The Comprehensive R Archive Network*, 337, 338.
- Rodriguez, C. I., Herman, D., Alcon, J., Chen, S., Tannen, A., Essock, S., & Simpson, H. B. (2012). Prevalence of hoarding disorder in individuals at potential risk of eviction in New York City: a pilot study. *The Journal of Nervous and Mental Disease*, 200(1).
- Saxena, S., Ayers, C. R., Maidment, K. M., Vapnik, T., Wetherell, J. L., & Bystritsky, A. (2011). Quality of life and functional impairment in compulsive hoarding. *Journal of Psychiatric Research*, 45(4), 475–480.
- Sousa, V. D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice*, 17(2), 268–274.
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models.
- Timpano, K. R., Broman-Fulks, J. J., Glaesmer, H., Exner, C., Rief, W., Olatunji, B. O., et al. (2013). A taxometric exploration of the latent structure of hoarding. *Psychological Assessment*, 25(1), 194.
- Timpano, K. R., Çek, D., Fu, Z.-F., Tang, T., Wang, J.-P., & Chasson, G. S. (2015). A consideration of hoarding disorder symptoms in China. *Comprehensive Psychiatry*, 57, 36–45.
- Tolin, D. F., Frost, R. O., & Steketee, G. (2010). A brief interview for assessing compulsive hoarding: the Hoarding Rating Scale-Interview. *Psychiatry Research*, 178(1), 147–152.
- Tolin, D. F., Frost, R. O., Steketee, G., & Fitch, K. E. (2008a). Family burden of compulsive hoarding: results of an internet survey. *Behaviour Research and Therapy*, 46(3), 334–344.
- Tolin, D. F., Frost, R. O., Steketee, G., Gray, K. D., & Fitch, K. E. (2008b). The economic and social burden of compulsive hoarding. *Psychiatry Research*, 160(2), 200–211.
- Tolin, D. F., Meunier, S. A., Frost, R. O., & Steketee, G. (2011). Hoarding among patients seeking treatment for anxiety disorders. *Journal of Anxiety Disorders*, 25(1), 43–48.
- Tortella-Feliu, M., Fullana, M. A., Caseras, X., Andión, Ó., Torrubia, R., & Mataix-Cols, D. (2006). Spanish version of the Savings Inventory–Revised: adaptation, psychometric properties, and relationship to personality variables. *Behavior Modification*, 30(5), 693–712.

Tsuchiyagaito, A., Horiuchi, S., Igarashi, T., Kawanori, Y., Hirano, Y., Yabe, H., & Nakagawa, A. (2017). Factor structure, reliability, and validity of the Japanese version of the Hoarding Rating Scale-Self-Report (HRS-SR-J). *Neuropsychiatric Disease and Treatment, 13*, 1235.

Walker, C. M. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development

and validation. *Journal of Psychoeducational Assessment, 29*(4), 364–376.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Caitlin A. Stamatis^{1,2}  · Jordana Muroff³  · Elizabeth S. Bocanegra⁴  · Carolyn I. Rodriguez^{5,6}  · Kiara R. Timpano¹ 

Caitlin A. Stamatis
caitlin.stamatis@gmail.com

Jordana Muroff
jmuroff@bu.edu

Elizabeth S. Bocanegra
ebocanegra@ucla.edu

Carolyn I. Rodriguez
carolynrodriguez@stanford.edu

¹ Department of Psychology, University of Miami, 5665 Ponce de Leon Blvd, Coral Gables, FL, USA

² Department of Psychiatry, New York-Presbyterian Hospital/Weill Cornell Medical College, New York, NY, USA

³ School of Social Work, Boston University, Boston, MA, USA

⁴ Department of Psychology, University of California, Los Angeles, Los Angeles, CA, USA

⁵ Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA

⁶ Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, USA