7-19-2013

Psychometric Properties of the Paper-and-Pencil And Online Versions of the Italian Saving Inventory-revised In Nonclinical Samples

Gabriele Melli
Institute of Behavioral and Cognitive Psychology and Psychotherapy

Carlo Chiorri
Università degli Studi di Genova

Rosa Smurra
Institute of Behavioral and Cognitive Psychology and Psychotherapy

Randy O. Frost
Smith College, rfrost@smith.edu

International Journal of Cognitive Therapy, 6(1), 40–56, 2013

Psychometric Properties of the Paper-and-Pencil and Online Versions of the Italian Saving Inventory–Revised in Nonclinical Samples

Gabriele Melli*
Institute of Behavioral and Cognitive Psychology and Psychotherapy, Florence, Italy

Carlo Chiorri*
University of Genoa, Italy

Rosa Smurra
Institute of Behavioral and Cognitive Psychology and Psychotherapy, Florence, Italy

Randy O. Frost
Smith College, Northampton, MA

Three studies were performed to investigate the psychometric properties of the paper-and-pencil and online versions of the Italian Saving Inventory-Revised (SI-R) in nonclinical participants. In Study 1, the SI-R was administered to a community sample of 473 participants together with measures of obsessive-compulsive symptomatology, compulsive shopping, depression, and anxiety. In Study 2, temporal stability of the SI-R was investigated by administering the scale to 75 participants twice with a 4-week interval in between. In Study 3, 452 participants completed the SI-R through the internet. Evidence of internal consistency, test-retest reliability, construct validity, and replicability of the original three-correlated-factor structure was obtained. After ruling out the bias due to nonrandomized assignment to administration methods through propensity matching, multiple-group confirmatory factor analyses provided evidence of measurement invariance among the administration formats. However, higher latent and manifest mean scores were observed in online participants, consistent with the "online disinhibition effect." These results suggest that the Italian version of the SI-R retains the psychometric properties of the original in both the paper-and-pencil and online versions, though different norms may be needed.

*The two authors contributed equally.

Address correspondence to Gabriele Melli, Ph.D., Via Mannelli, 139 - 50132 Firenze, Italy. E-mail: g.melli@ipsico.it.

Hoarding is characterized by the acquisition of, and failure to discard, a large number of possessions that cover the living areas of the home and cause significant distress or impairment (Frost & Hartl, 1996). Hoarding is associated with substantial disability and functional impairment in daily life (Saxena, Ayers, Maidment et al., 2011; Tolin, Frost, Steketee, Gray, & Fitch, 2008), and with increased risks of fire, falling, and illness (Kim, Steketee, & Frost, 2001). While most studies published in the last two decades have considered hoarding as a symptom dimension of OCD, there is a growing body of evidence supporting the disorder as a separate diagnosis. Because of its distinctiveness and potentially serious consequences, hoarding has been recommended for inclusion in *DSM-V* as a separate disorder (Mataix-Cols, Frost, Pertusa et al., 2010).

Recent research has focused on developing measures of hoarding (for a review, see Frost & Hristova, 2011). The most widely used self-report measure of hoarding is the Saving Inventory–Revised (SI-R; Frost, Steketee, & Grisham, 2004). The SI-R is a 23-item questionnaire consisting of three subscales designed to measure the prominent features of hoarding disorder: excessive clutter in the home, difficulty discarding possessions, and excessive acquisition of purchased and free items. Items were designed to measure distress and impairment associated with each of these dimensions, to minimize the assessment of beliefs associated with hoarding which can be assessed separately (Steketee, Frost, & Kyrios, 2003), and to avoid references to specific types of possessions, since these vary from patient to patient and cannot be applied universally (Frost et al., 2004). Participants are asked to rate the extent to which each statement describes them on a 5-point, Likert-type scale ranging from 0 (strongly disagree) to 4 (strongly agree). The English language version of the SI-R has a robust three-factor solution, adequate internal consistency for the total score and the three subscales, and adequate test-retest reliability for the total score and subscales. The convergent validity of the SI-R was supported by strong correlations with other indices of hoarding including observational, interview, and self-report measures (see Frost & Hristova, 2011). Discriminant validity was demonstrated by relatively weaker correlations with non-hoarding constructs (e.g., anxiety, negative affect, etc.). The SI-R distinguishes hoarding from OCD participants without hoarding as well as from community controls. It is also sensitive to cognitive behavior therapy treatment effects (Steketee, Frost, Tolin, Rasmussen, & Brown, 2010).

The SI-R has been adapted and validated in a number of other languages including German (Mueller, Crosby, Frost et al., 2009), Spanish (Tortella-Feliu, Fullana, Caseras et al., 2006), Portuguese (Fontenelle, Prazeres, Borges et al., 2010), and Farsi (Mohammadzadeh, 2009). The purpose of this investigation was to develop an Italian version of the SI-R and to examine its psychometric properties in nonclinical samples. Three studies were performed. Study 1 examined internal consistency, factor structure, and construct validity of the SI-R. Study 2 tested the temporal stability and consistency of the SI-R, and Study 3 examined the measurement equivalence of the SI-R across online and paper-and-pencil formats. Test and questionnaire administration via the Internet has been gaining popularity, and research has shown both its benefits and disadvantages (for a comprehensive

review, see Gosling & Johnson, 2010). However, it has also been suggested that measurement equivalence across online and paper-and-pencil formats of the same scale cannot be taken for granted, and needs to be supported by empirical evidence (e.g., Buchanan, Ali, Heffernan et al., 2005).

The participants in the three studies were independent samples. All studies were approved by the Institutional Review Board of the University of Padova and participants were treated in accordance with the *Ethical Principles of Psychologists and Code of Conduct* (American Psychological Association, 2002).

STUDY 1

Methods

Translation of the SI-R into Italian

The Italian version of the SI-R was developed through a mixed forward- and back-translation procedure (Behling & Law, 2000). Two of the authors and one bilingual Italian-English psychologist independently translated the English version of the SI-R into Italian. After consensus among translators was achieved, an Italian-English person, blind to the original version, translated this preliminary version back into English. Discrepancies were discussed among the translators until an agreement on a common version was reached. The newly developed Italian version of the SI-R was administered to ten naive participants in order to check understandability of the items. All items were found to be easy to understand and score.

Participants

Responding to advertisements requesting potential volunteers for psychological studies in North-Central Italy, 480 people volunteered to participate following a detailed description of the procedures. Participants had to be 18 years of age or older, possess at least a primary school education, and had not been treated for any psychiatric disorder. Seven participants reported that they received an Axis I diagnosis and were excluded. The final sample consisted of 473 (53.9% female) community volunteers with a mean age of 38.78 (range = 18–84; *SD* = 14.55). Based on the Italian education system, 45.7% of the participants had a medium level of education (12–13 years, high school degree), 37.4% had a high level (16 or more years, degree or Ph.D.), and the remaining 16.9% had a low level (8 or less years, primary or secondary school license). Most were employed (n = 265; 56.0%), 125 (26.4%) were undergraduate university students, and the remaining 17.6% were housewives, unemployed, or retired. Most were single (n = 230; 48.6%), while 204 (43.1%) were married or cohabiting, 26 (5.5%) were divorced, and 13 (2.7%) were widows or widowers.

Measures

Measures of Hoarding

Saving Inventory-Revised (SI-R; Frost et al., 2004). As described in the Introduction.

Measures of Obsessive-Compulsive Behavior

Vancouver Obsessive Compulsive Inventory (VOCI). The VOCI (Thordarson, Radomsky, Rachman, et al., 2004) is a 55-item self-report questionnaire that assesses a variety of symptoms and characteristics of obsessive-compulsive disorder (OCD). Along with a total score, the VOCI also contains six subscales, each assessing a specific symptom area of OCD: Contamination, Checking, Obsessions, Hoarding, Just Rightness, and Indecisiveness. The VOCI exhibited good internal consistency and good test-retest reliability within both the student and the OCD samples. The VOCI also showed good convergent and divergent validity in both clinical and student populations. The Italian version of the VOCI (Chiorri, Melli, & Smurra, 2011) has shown adequate psychometric properties. Internal consistency ranged from .78 to .89, whereas test-retest correlations ranged from .51 to .80. Correlations of the Italian VOCI with other measures of OCD (e.g., Padua Inventory and the Y-BOCS) supported its construct validity.

Compulsive Buying Measurement Scale (CBMS). The CBMS (Valence, d'Astous, & Fortier, 1988) is a 13-item self-report measure for the assessment of buying addiction. Items relate to difficulties in: managing money, impulsive buying behavior, the use of buying as a coping mechanism, and the feeling of guilt following buying binges. In a series of studies, the Italian version of the CBMS (Pani & Biolcati, 1998, 2006) has been found to have adequate internal consistency and correlate with other measures of compulsive buying (e.g., Compulsive Buying Scale; Edwards, 1993).

Measures of Depression and Anxiety

Beck Depression Inventory-II (BDI-II). The BDI-II (Beck, Steer, & Brown, 1996) consists of 21 groups of four alternative self-evaluative statements referring to affective, cognitive, motivational, psychomotor, and vegetative components of depression. Studies of the Italian version of the BDI-II (Ghisi, Flebus, Montano, Sanavio, & Sica, 2006; Sica & Ghisi, 2007) reported adequate internal consistency (αs in the range .80-.87), test-retest reliability (r = .76) and construct validity.

State-Trait Anxiety Inventory-Trait (STAI-T). The STAI-T (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) is a 20-item self-report measure of anxiety proneness requiring participants to rate their frequency of anxiety symptoms on a 4-point Likert frequency scale. The Italian version of the STAI-T (Sanavio, Bertolotti, Michielin, Vidotto, & Zotti, 1997) has been found to have adequate internal consistency ($\alpha \geq .90$) and test-retest reliability (rs = .73 to .86).

Procedure

All participants were tested individually in a single session on the premises either of a psychotherapy institute in Central Italy or a university in Northern Italy. The scales were administered in a counterbalanced fashion to control for order and sequence effects.

Data Analysis

Data analyses were performed using SPSS 13.0 (SPSS Inc., 2006), MPLUS 6.1 (Muthén & Muthén, 1998-2010), and FACTOR8 (Lorenzo-Seva & Ferrando, 2006). The factor structure of the Italian paper-and-pencil SI-R was assessed by conducting Exploratory Factor Analyses (EFAs). Since univariate skewness and kurtosis were found to be indicative of substantial non-normality through preliminary distribution analyses, EFAs were performed starting from a polychoric correlation matrix using a robust weighted least square estimator (WLSMV) which used a diagonal weight matrix with standard errors and a mean- and variance-adjusted chi-square test statistic that used a full-weight matrix (Muthén, du Toit, & Spisic, 1997). This provided the Root Mean Square Error of Approximation (RMSEA) as a fit index which indicates acceptable and optimal fits when smaller than .08 and .06, respectively (Marsh, 2007).

Internal reliability coefficients (Cronbach's αs), and correlations of SI-R scores with demographic and other variables were computed. We used the $Z_{contrast}$ test (Westen & Rosenthal, 2003) to compare the SI-R–VOCI Hoarding correlations with the correlations of SI-R scores with non-corresponding constructs.

Results

Exploratory Factor Analyses

Parallel analysis (PA) and scree plot (SC) suggested a three-factor solution (First ten observed eigenvalues: 11.94, 1.52, 1.45, 0.99, 0.81, 0.76, 0.74, 0.62, 0.58, 0.53; 95th percentile of random eigenvalues: 8.00, 1.50, 1.38, 1.29, 1.22, 1.15, 1.08, 1.04, 0.97, 0.92), while the Minimum Average Partial Correlation Statistic (MAP) values (.0217, .0158, .0280, .0625, .1495, .7039, 1.000) indicated an optimal solution with only two factors. A two-correlated-factor solution grouped all Difficulty Discarding items in one latent dimension and Acquisition and Clutter items in the other. Though an apparent simple structure was obtained, RMSEA was .101, suggesting sub-optimal fit.

Conversely, a three-correlated-factor solution provided a simple structure, and RMSEA (.063) was in the acceptable range. As shown in Table 1, all items loaded on the expected factor; the highest cross-loading was .35 for item 22, but it was substantially lower than the target loading (.45).

TABLE 1. Results of WLSMV Factor Analyses of the Saving Inventory-Revised (SI-R) for Study 1 Participants

		Three-factor solution			
SI-R Item Content	**Scale**	**Factor 1**	**Factor 2**	**Factor 3**	h^2
2. how distressing is discarding	Difficulty Discarding	**.91**	.02	–.06	.78
1. extent of difficulty throwing things away	Difficulty Discarding	**.91**	–.01	–.04	.77
4. avoid discarding because of distress	Difficulty Discarding	**.84**	.02	.01	.74
23. inability to Difficulty Discarding	Difficulty Discarding	**.66**	.13	.14	.69
16. strength of urge to save	Difficulty Discarding	**.61**	.23	.12	.71
11. frequency of saving things not needed	Difficulty Discarding	**.52**	.22	.18	.63
17. lack of control over urges to save	Difficulty Discarding	**.49**	.04	.06	.30
9. strength of urge to acquire	Acquisition	.00	**.82**	.10	.77
8. feel compelled to acquire	Acquisition	.07	**.79**	-.11	.61
19. distress over acquiring habits	Acquisition	–.09	**.76**	.05	.56
15. frequency of buying unneeded items	Acquisition	.06	**.68**	.14	.65
5. distress if can't acquire	Acquisition	.23	**.67**	–.11	.57
10. lack of control over urges to acquire	Acquisition	–.01	**.49**	.04	.26
22. financial difficulties from saving/buying	Acquisition	–.03	**.45**	**.35**	.47
7. clutter interferes with functioning	Clutter	–.16	.11	**.89**	.76
18. clutter restricts navigation through home	Clutter	.00	–.03	**.82**	.65
3. extent of clutter	Clutter	.10	.00	**.79**	.74
13. clutter causes distress	Clutter	.00	.14	**.78**	.75
14. clutter prevents inviting people over	Clutter	-.09	.22	**.76**	.71
20. clutter prevents normal activities	Clutter	.04	.12	**.75**	.72
12. clutter prevents use of home	Clutter	.24	–.29	**.68**	.52
6. clutter in living areas	Clutter	.15	–.08	**.57**	.40
21. lack of control over clutter	Clutter	.11	.21	**.54**	.57
	Correlation with Factor 2	**.55**			
	Correlation with Factor 3	**.61**	**.54**		

Note. h^2 = item communalities. When the WLSMV method is used, communalities are not the simple sum of squared loadings but MPlus provides estimates of residual variances, hence communality estimates can be derived as 1-residual variance. Factor loadings and correlations higher than |.30| are bolded for ease of interpretation.

Item Analysis

Cronbach's αs were higher than .80 for all SI-R scales (Table 2). Corrected item-total correlations were all higher than .20 (Nunnally & Bernstein, 1994) and averaged higher than .55 (after *r*-to-*Z* transformation and back-transformation). Total and scale scores were all significantly and strongly correlated (i.e., *r*s > .50). Mean scores (standard deviations) for the sample were 16.99 (11.83) for SI-R Total Score, 5.66 (4.27) for Acquisition, 4.84 (4.93) for Clutter, and 6.48 (4.87) for Difficulty Discarding. The SI-R subscales were strongly intercorrelated (*r*s from .52 to .59, Table 2).

TABLE 2. Correlation Matrix of SI-R Total and Subscale Scores with Other Constructs (Pearson's *rs*)

Scale		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.	SI-R - Total Score	*.92*													
2.	SI-R - Acquisition	**.75°**	*.81*												
3.	SI-R - Clutter	**.81°**	**.52**	*.86*											
4.	SI-R - Discard	**.81°**	**.57**	**.59**	*.88*										
5.	VOCI - Total Score	**.54**	**.46**	**.49**	**.42**	*.94*									
6.	VOCI - Checking	.27	.28	.26	.16	**.79°**	*.84*								
7.	VOCI - Contamination	**.37**	**.34**	**.32**	.29	**.81°**	**.63**	*.86*							
8.	VOCI - Hoarding	**.65**	**.38**	**.60**	**.64**	**.75°**	**.35**	**.48**	*.80*						
9.	VOCI - Indecisiveness	**.43**	**.35**	**.37**	**.37**	**.67°**	**.40**	**.48**	**.48**	*.71*					
10.	VOCI - Just Right	**.48**	**.42**	**.44**	**.35**	**.77°**	**.58**	**.66**	**.57**	**.61**	*.80*				
11.	VOCI - Obsession	**.37**	**.38**	**.33**	.22	**.82°**	**.50**	**.49**	**.40**	**.51**	**.59**	*.87*			
12.	CBMS - Total Score	**.48**	**.62**	.27	**.34**	**.44**	.29	**.41**	.26	.28	**.36**	**.37**	*.92*		
13.	STAI-T - Total Score	**.39**	**.34**	**.35**	**.30**	**.50**	.26	**.30**	**.30**	**.61**	**.44**	**.45**	**.31**	*.90*	
14.	BDI - Total Score	.29	.24	.27	.21	**.41**	.27	.28	.27	**.40**	**.33**	**.40**	.24	**.63**	*.90*

Note. SI-R = Saving Inventory Revised; VOCI = Vancouver Obsessive Compulsive Inventory; CBMS = Compulsive Buying Measurement Scale; STAI-T: State-Trait Anxiety Inventory Trait; BDI = Beck Depression Inventory.

° = coefficient corrected for common items. Since the total score includes the subscale scores as components, a correlation between the total score and each subscale will be artificially inflated because both the subscale and the total score contain the subscale's variance. The correction formula suggested by Nunnally and Bernstein (1994) was applied.

Italicized coefficients on the main diagonal represent scale internal consistencies (Cronbach's alphas). All correlations higher than |.09| were significant at a .05 (not corrected for multiple tests). Correlation coefficients higher than |.30| are bolded for ease of interpretation.

Associations with Demographic Variables

Women scored higher than men in all scales: Total Score (*F:* 19.43±12.26 vs. *M:* 14.12±10.66, $t(471) = 4.97, p < .001, r = .22$), Acquisition (*F:* 6.59±4.36 vs. *M:* 4.58±3.91, $t(471) = 5.23, p < .001, r = .24$), Clutter (*F:* 5.55±5.59 vs. *M:* 4.01±3.87, $t(471) = 3.42, p < .001, r = .16$), and Difficulty Discarding (*F:* 7.29±5.01 vs. *M:* 5.54±4.54, $t(471) = 3.96, p < .001, r = .18$). Acquisition was the only scale significantly correlated with age, though weakly ($r = -.20, p < .01$), while no correlation (Spearman's *rho*) was found with education. ANOVA tests revealed significant differences (corrected $p < .013$) among groups defined by marital status in Total Score ($F(3, 469) = 4.68, p = .003, r = .17$), and Acquisition ($F(3, 469) = 8.35, p < .001, r = .23$). Post-hoc tests revealed that, in both cases, widows and widowers obtained lower scores than singles who, in their turn, scored higher than married/cohabiting on the Total Score. ANOVA performed using the occupation as grouping variable revealed that students scored higher than office workers in Clutter ($F(6, 466) = 3.52, p = .002, r = .21$).

Construct Validity of the Italian SI-R

The correlations of SI-R Total Score, Clutter, and Difficulty Discarding scores with VOCI-Hoarding were greater in magnitude than those with any other scale (i.e., the other VOCI subscales, CBMS, BDI-II, and STAI-T; Table 2; $Z_{contrast} = 8.94, p < .001, r = .40$; $Z_{contrast} = 8.67, p < .001, r = .39$; $Z_{contrast} = 11.49, p < .001, r = .50$, respectively). In contrast, the SI-R Acquisition subscale showed the highest correlation with CBMS ($Z_{contrast} = 9.18, p < .001, r = .41$), and this correlation was also statistically higher than the correlation of Acquisition with VOCI-Hoarding ($Z_{contrast} = 6.09, p < .001, r = .28$). Notably, the correlation of the Acquisition subscale with VOCI-Hoarding was not substantially higher than its correlation with anxiety, depression, and the other VOCI subscales.

STUDY 2

Methods

Participants

Recruitment and inclusion/exclusion criteria were identical to Study 1. Seventy-five participants (44 females) took part in this study with a mean age of 45.75 years (*SD* = 11.87). Based on the Italian education system, 38.7% of participants had a medium level of education, 37.3% had a high level of education, and the remaining 24.0% had a low level of education. Sixty-five (86.7%) participants were employed, six (8.0%) were undergraduate university students, while the remaining 6.3% were housewives, unemployed, or retired. Twenty-one (28.0%) were single, 47 (62.7%) were married or cohabiting, four (5.3%) were divorced, and

three (4.0%) were widows or widowers. All participants volunteered to take part after being presented with a detailed description of the procedure.

Procedure

The paper-and-pencil version of the SI-R was administered twice with a 4-week interval on the premises of a university in Northern Italy. Each administration took place on the same day for all participants and took approximately 10 minutes to complete.

Results

Test-retest correlations (r_{tt}) and intraclass correlation coefficients (ICC) were computed using a two-factor-mixed effect model and type consistency (McGraw & Wong, 1996). The findings indicated that SI-R scores were stable over the 4-week interval. Test-retest reliabilities for the SI-R–Total Score (r_{tt} = .88; ICC = .94), Acquisition (r_{tt} = .83; ICC = .90), Clutter (r_{tt} = .79; ICC = .83), and Difficulty Discarding (r_{tt} = .91; ICC = .95) were all high. Paired *t*-tests indicated no significant mean changes from Time 1 to Time 2.

STUDY 3

Participants

Potential participants included 589 people who visited an interactive website created on the internet. Participants were recruited through authors' and their assistants' e-mail address contacts. They received an e-mail invitation that included a short description of the study and an access token randomly generated by the software. To gain access to the website, they had to click their unique study identification link. To maintain anonymity, authors did not know participants' identification links. Once logged in, participants were presented with a detailed description of the procedure. To be allowed to answer the inventory, they had to answer "Yes" to a question asking whether they accepted participation in the study.

Inclusion criteria were the same as in Study 1. Of the 589 potential participants, 64 were excluded because they never connected to the website, 63 were excluded because they did not provide any answers, and 10 because they did not complete the full inventory. The final sample consisted of 452 (76.3% female) participants. None of these reported having received a psychiatric diagnosis. Ages ranged from 18 to 70 with a mean of 33.75 (*SD* = 9.13). Based on the Italian education system, 28.3% of participants had a medium level of education, 69.1% had a high level of education, and 2.7% had a low level of education. Regarding employment, 361 participants (79.9%) were employed, 63 (13.9%) were under-

graduate university students, and 6.2% were housewives, unemployed, or retired. Most were single (n = 265, 58.6%), 160 (35.4%) were married or cohabiting, 24 (5.3%) were divorced, and three (0.7%) were widows or widowers.

Materials and Procedure

Participants were given the demographic questions, identical to Study 1, and the SI-R. The questions were presented on separate full-page layouts such that participants could view each page of the questionnaire by scrolling up or down. They were not required to answer all items.

Data Analysis

Data analyses were performed using SPSS 13.0, Mplus 6.1, FACTOR8, and the MatchIt package (Ho, Imai, King, & Stuart, 2011; Ho, Stuart, Imai, & King, 2011) in R (R Development Core Team, 2009).

Six measurement models were tested using Confirmatory Factor Analysis (CFA). The first model (Model 1, M1) specified that all items loaded on a single, first-order Saving factor. Model 2 (M2) and Model 3 (M3) identified the two factors (Difficulty Discarding vs. Clutter-Acquisition) found in the Study 1 EFA, and specified that they be independent and correlated, respectively. Model 4 (M4) and Model 5 (M5) specified the three factors from the original EFA (Acquisition, Clutter, and Difficulty Discarding) and required that these three factors were independent and correlated, respectively. Since distribution analyses showed a substantial non-normality of item distributions, a WLSMV estimation was used.

One of the aims of this article was also to test whether SI-R mean scores would differ across administration methods, that is, to what extent the *net difference* observed between scores on the paper-and-pencil and online participants can be attributed to the administration method, all other variables kept constant. Since we could not randomize the assignment of participants to the administration method, a clear sampling bias was introduced, and, in fact, the distribution of covariates across the paper-and-pencil and online samples was unbalanced (see the Results section). To address this issue and obtain unbiased estimates of mean score differences across administration methods, Propensity Score Analysis (PSA; Rosenbaum & Rubin, 1983) was used.[1] PSA is a matching technique employing propensity scores, that is, the conditional probability of assignment to a treatment condition, given a vector of observed covariates, to create a counterfactual group. The matching procedure allowed us to obtain weights for each participant in the

1. In this case, an analysis of covariance (ANCOVA) approach is doomed to provide biased and inconsistent estimates of group differences unless all relevant interaction and nonlinear effects are included in the model (Schafer & Kang, 2008).

TABLE 3. Summary of Goodness-of-Fit Statistics for the Confirmatory Factor Analysis Models and the Multiple-Group Confirmatory Factor Analysis Models Testing Invariance Across Administration Formats and Patterns of Differences on Saving Facet Latent Mean Scores (Standardized Coefficients)

Model	χ^2	*df*	*CFI*	*TLI*	*RMSEA*
	Confirmatory Factor Analysis Models				
M1 Single hoarding factor	2510.12	230	.82	.80	.148
M2 Two uncorrelated factors (Difficulty Discarding and Clutter+Acquisition)	5144.75	230	.61	.57	.217
M3 Two correlated factors (Difficulty Discarding and Clutter+Acquisition)	1678.82	229	.89	.87	.118
M4 Three uncorrelated factors	4845.51	230	.64	.60	.211
M5 Three correlated factors	812.38	227	.95	.97	.076
	Multiple Group Confirmatory Factor Analysis Models				
MG1 configural invariance	921.59	457	.97	.96	.047
MG2 FL weak invariance	937.66	477	.97	.97	.046
MG3 FL TH strong invariance	968.43	517	.97	.97	.043
MG4 FL TH Uniq strict invariance	1008.08	540	.97	.97	.043
MG5 FL TH Uniq FV FCov	956.63	546	.97	.97	.040
MG6 FL TH Uniq FV FCov LFM complete invariance	1080.51	549	.96	.97	.046
	Standardized Latent Mean Score Differences°				
	Acquisition	Clutter	Difficulty Discarding		
MG3	.185	.586***	.715***		
MG4	.166	.570***	.739***		
MG5	.129	.522***	.754***		

Note. FL = factor loadings; TH = thresholds; Uniq = uniquenesses (residual variances); FV = factor variances; FCov = Factor Covariances; LFM = Latent factor means. *** = $p < .001$; ° = positive coefficients indicate higher scores in the online sample.

paper-and-pencil group, which was the "control" group, and obtain an adequate balance of the covariates (i.e., $r < .25$ on each covariate, Guo & Fraser, 2010) with respect to the online group, which was the "treatment" group. It has been convincingly shown that PSA can effectively rule out the bias in the estimate of mean differences due to non-random assignment in observational studies (e.g., Dehejia & Wahba, 1999). We then assessed the SI-R measurement invariance across different administration methods through multiple-group confirmatory factor analysis (MG-CFA) on matched data.

MG-CFA allows for the examination of the equivalence (i.e., invariance) of the measurement and structural models across multiple groups, that is, whether the scales measure the same constructs in the same way and whether the measurements themselves are operating in the same way across groups. If this were not the case, mean differences and other comparisons would likely be invalid (e.g., Brown, 2006). In imposing equality constraints across groups on model parameters we followed Millsap and Tein's (2004) recommendations for the assessment of factorial invariance with ordered-categorical items. We tested whether: (1) the *a priori* model fit the data in each group (*configural invariance model*, MG1, no equality constraints); (2) factor loadings were equal across groups (*weak in-*

variance model, MG2); (3) factor loadings and item thresholds were equal across groups (*strong invariance model*, MG3); (4) factor loadings, item thresholds, and item residual variances were equal across groups (*strict invariance model*, MG4); (5) factor loadings, item thresholds, item residual variances, and factor variances/covariances were equal across groups (MG5); (6) factor loadings, item thresholds, item residual variances, factor variances/covariances, and latent means were equal across groups (*complete invariance model,* MG6).

The two highest response categories on the SI-R were rarely endorsed, which resulted in item distributions that were positively skewed. These response categories were combined and resulted in four response categories for the SI-R items.

In both the CFA and MG-CFAs, model fit was assessed through the Correct χ^2 Statistic (S-B χ^2), the Tucker-Lewis Index (TLI), and the Comparative Fit Index (CFI; acceptable fit for both indices: CFI $\geq$ .90; Marsh, 2007), and the RMSEA (acceptable fit: RMSEA $\leq$ .08). Model comparison was not based on a chi-square difference test since its significance is heavily dependent on sample size, but we considered it as supporting evidence for a more parsimonious model a change in CFI of less than .01 (Chen, 2007) or a change in RMSEA of less than .015 (Chen, 2007).

Results

Confirmatory Factor Analyses

As shown in Table 3 the three-correlated-factor model showed the best fit to the data. Other models' goodness-of-fit indices revealed substantial lack of fit. It can be argued that these results do not fully rule out the possibility of other factor solutions providing a better fit for the data in this study. We therefore performed dimensionality analyses on Study 1 data, and, as in Study 1, we found that two (PA and SP) out of three analyses of the item pool dimensionality suggested a three-factor solution, whereas MAP indicated an optimal solution with only two factors. A two-factor WLSMV-EFA with promax rotation grouped together Difficulty Discarding, and Clutter items in one factor, and Acquisition items in the other. This solution (not reported here, but results are available from the corresponding author) showed three items (2, 14, and 20) with cross loadings higher than |.30| but substantially lower (difference > |.10|) than target loadings, and one item (7) with loadings of .45 and .37 on the two factors, respectively. In addition, RMSEA was .112. Conversely, a three-factor solution yielded a simple structure with all items loading on the *a priori* expected factor and acceptable indices of lack of fit (RMSEA = .065). Taken together, these results suggest that it is unlikely that factor solutions other than the expected three-correlated-factor solution could provide a better fit for the data in this study.

Item Analysis

Cronbach αs were higher than .80 for all SI-R scales (Total Score: .93, Acquisition: .84, Clutter: .89, Difficulty Discarding: .88), corrected item-total correlations

were all higher than .20, averaging higher than .55 (after *r*-to-*Z* transformation and back-transformation). Subscale scores were all strongly intercorrelated (*r*s from .48 to .63). Mean scores (standard deviations) were 24.41 (12.75) for SI-R Total Score, 7.13 (4.35) for Acquisition, 7.14 (5.51) for Clutter, and 10.15 (5.41) for Difficulty Discarding.

Associations with Demographic Variables

SI-R scores did not show any significant gender difference or association with age. Acquisition was weakly associated with educational level (Spearman's *rho* = -.26, $p < .001$). SI-R scores did not differ with respect to marital status. After correction of the comparison-wise significance level for multiple comparisons, significant differences were found among occupation categories in: Total Score, $F(6, 451) = 3.62$, $p = .002$, $r = .22$, Acquisition, $F(6, 451) = 6.07$, $p < .001$, $r = .27$, and Difficulty Discarding, $F(6, 451) = 3.95$, $p = .006$, $r = .19$. In all these cases, Games-Howell post-hoc tests revealed that professionals scored lower than students and office workers.

Pattern of Differences in Covariates among Paper-and-Pencil and Online Participants

Compared to online participants in this study, the paper-and-pencil participants from Study 1 were more likely to be males ($X^2(1) = 50.96$, $p < .001$, $r = .23$), were older ($t(923) = 6.26$, $p < .001$, $r = .20$), had fewer years of education ($t(923) = -12.65$, $p < .001$, $r = .38$), were less likely to be single and more likely to be married or widowed ($X^2(3) = 13.65$, $p = .003$, $r = .12$),[2] were more likely to be housewives, clerks, retired, and students and less likely to be professionals ($X^2(6) = 122.91$, $p < .001$, $r = .37$). These results support the need for a PSA to obtain unbiased estimates of scale scores differences.

Measurement Invariance

All MG-CFA models tested using matched data showed a good fit (Table 3). However, while adding constraints did not produce substantial differences in fit across the first five models, a decrease in fit was observed when latent means were constrained to equality across groups. This result suggested that latent means could be significantly different across groups, as witnessed by different coefficients in the bottom part of Table 3. Positive coefficients indicate higher scores in the online group which were significant for Clutter and Difficulty Discarding.

Differences in Observed Scores

Though MG4 already allowed for a formal test of manifest mean score difference, we performed a more traditional regression-like analysis on matched sample observed scores to simultaneously test the effect of all demographic variables on

2. Results are drawn from the inspection of adjusted standardized residuals of the chi-square test for independence of categorical variables.

scale scores after the differences in administration methods have been ruled out. We could not perform a multivariate analysis of covariance (ANCOVA) since variance-covariance matrices were not equal across groups (Box's $M = 71.96$, $F(18, 455507.4) = 3.96$, $p < .001$). Therefore, we ran separate ANCOVAs on each scale score using the administration method as focal variable and specifying the main effect of all other variables and gender by administration method interaction. The model with Total Score as criterion showed higher scores being associated with Internet administration (estimated marginal means [EMMs; standard error, *SE*]; Paper-and-pencil: 17.74 [0.71], Online: 24.48 [0.67]; $t = 4.64$, $p < .001$, $r = .22$) and a positive association with age ($r = .07$, $t = 2.07$, $p = .039$). Acquisition was not significantly associated with any variable. Clutter was also significantly higher with internet administration (EMMs [SE]; Paper-and-pencil: 4.84 [0.29], Online: 7.27 [0.28], $t = 6.76$, $p < .001$, $r = .31$) and positively correlated with age ($r = .09$, $t = 2.84$, $p = .005$). Difficulty Discarding was higher for internet administration as well (EMMs [SE]; Paper-and-pencil: 6.50 [0.30], Online: 10.17 [0.28], $t = 8.59$, $p < .001$, $r = .38$), was positively correlated with age ($r = .10$, $t = 3.06$, $p = .002$), and showed a gender by administration method interaction ($t = 2.15$, $p = .032$, $r = .07$). This effect was due to higher scores of females in the paper-and-pencil group (EMMs [SE]; Females: 7.08 [0.27], Males: 5.91 [0.53]), whereas the gender group means did not differ in the online group (EMMs [SE]; Females: 9.87 [0.27], Males: 10.47 [0.49]).

DISCUSSION

The findings reported in this article suggest that in nonclinical samples the Italian version of the SI-R shows sound psychometric properties with respect to factor structure, internal consistency of scales, and temporal stability of scores. Evidence of convergent and discriminant validity were also obtained since the SI-R Total Score, Clutter, and Difficulty Discarding correlated more strongly with measures of hoarding than with measures of OCD, depression, and anxiety. The SI-R Acquisition subscale showed a different pattern of correlations. It was most highly correlated with compulsive buying (CBMS), which might be expected since the constructs are similar. In contrast to the other SI-R subscales, Acquisition did not correlate more strongly with the VOCI Hoarding subscale than other VOCI subscales, nor was this correlation substantially higher than those with measures of anxiety or depression. However, the VOCI Hoarding subscale includes only items having to do with clutter and difficulty discarding and it does not contain any item reflecting the acquisition of possessions. This probably reflects more on the weakness of the VOCI Hoarding subscale, since excessive acquisition is a core feature of hoarding (Frost, Tolin, Steketee, Fitch, & Selbo-Bruns, 2009).

The results of this study also support the comparability of online and paper-and-pencil administration of the SI-R. The two versions had similar internal consistencies, mean corrected item-total correlations, and correlations among scale scores. EFAs and CFAs showed that the three-correlated-factor solution was

replicable across administration methods, even if the samples were quite different in demographic characteristics. MG-CFA supported strict measurement invariance for the three-factor models (MG4 and MG5 in Table 3). However, the mean scores differed across administration method groups. Online participants had higher scores on Clutter, Difficulty Discarding, and SI-R Total Score. Had we not employed the propensity score analysis (PSA), the demographic differences between the paper-and-pencil and online samples could have accounted for these differences. However, in observational studies like this one in which participants are not assigned randomly to conditions and the groups to be compared are unbalanced on a set of covariates, the PSA is known to yield estimates of mean score differences that are much closer to those that one would have obtained in a randomized controlled study (Dehejia & Wahba, 1999). Hence, these results could be explained in terms of the so-called "online disinhibition effect" proposed by Suler (2004), that is, a sort of loosening of social restrictions and inhibitions that would otherwise be present in face-to-face, and even paper-and-pencil, administration. Online participants might have become less guarded about expressing their true feelings and reporting behaviors they considered undesirable. These results suggest the use of different norms for the paper-and-pencil and the online versions of the Italian SI-R, albeit it should be noted that this study could not determine whether online participants were more accurate or over-reported clutter and difficulty discarding.

Substantial gender differences (women scoring higher than men) were found in the paper-and-pencil group, but not in the online one. Similarly, association of SI-R scores with other demographic characteristics showed no consistent pattern across groups. Students tended to report higher levels of hoarding than other occupational categories, which is consistent with the stereotypical image of an untidy and messy college student's room. However, when all these variables were entered in a single general linear model evaluating the effect of each variable while controlling for all the others, only a weak positive effect for age was found for all scores except Acquisition. Though cultural differences might also explain these results, no definite conclusions can be drawn as to the association of SI-R scores with demographic variables.

There are several limitations in the current studies. All three samples were nonclinical, which limits the generalizations that can be drawn from the findings. Replication using clinical samples is thus needed. Moreover, a further limitation to the generalizability of the results might have been introduced by the convenience sampling. Finally, the use of the VOCI Hoarding subscale to validate the SI-R may not have been the most optimal choice. The VOCI Hoarding items do not cover acquisition, a major feature in the majority of hoarding cases (Frost et al., 2009). Other well-validated measures of hoarding exist (See Frost & Hristova, 2011), but unfortunately they have yet to be adapted into Italian.

In conclusion, this investigation provided evidence that the Italian online and paper-and-pencil versions of the SI-R assess the same constructs in the same way in nonclinical participants, and that both can be confidently administered—provided that different norms are used. Moreover, the SI-R appears to be a reliable

and valid instrument for the assessment of hoarding behaviors and can be used in cultural contexts different from the original.

REFERENCES

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57*, 1060-1073.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II (BDI-II)*. San Antonio, TX: Psychological Corporation.

Behling, O., & Law, K. S. (2000). *Translating questionnaires and other research instruments: Problems and solutions* (Sage University papers series on quantitative applications in the social sciences, series no. 07-131). Thousand Oaks, CA: Sage.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.

Buchanan, T., Ali, T., Heffernan, T. M., Ling, J., Parrott, A. C., Rodgers, J., & Scholey, A. B. (2005). Nonequivalence of on-line and paper-and-pencil psychological tests: The case of the prospective memory questionnaire. *Behavior Research Methods, 37*, 148-154.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464-504.

Chiorri, C., Melli, G., & Smurra, R. (2011). Second-order factor structure of the Vancouver Obsessive Compulsive Inventory (VOCI) in a nonclinical sample. *Behavioral and Cognitive Psychotherapy, 39*, 561-577.

Dehejia, R. H., & Wahba, S. (1999). Causal effects in non-experimental studies: Re-evaluating the evaluation of training programs. *Journal of the American Statistical Association, 94*, 1053-1062.

Edwards, E. A. (1993). Development of a new scale for measuring compulsive buying behavior. *Financial Counseling and Planning, 4*, 67-84.

Fontenelle, I., Prazeres, A., Borges, M., Rangé, B., Versiani, M., & Fontenelle, L. (2010). The Brazilian Portuguese version of the Saving Inventory-Revised (SI-R): Internal consistency, test-retest reliability, and validity of a questionnaire to assess hoarding. *Psychological Reports*, *106*, 279-296.

Frost, R. O., & Hartl, T. (1996). A cognitive-behavioral model of compulsive hoarding. *Behaviour Research and Therapy*, *34*, 341-350.

Frost, R. O., & Hristova, V. (2011). The assessment of hoarding. *Journal of Clinical Psychology: In Session, 67,* 456-466.

Frost, R. O., Steketee, G., & Grisham, J. (2004). Measurement of compulsive hoarding: Saving Inventory-Revised (SI-R). *Behaviour Research and Therapy*, *42*, 1163-1182.

Frost, R. O., Tolin, D. F., Steketee, G., Fitch, K. E., & Selbo-Bruns, A. (2009). Excessive acquisition in hoarding. *Journal of Anxiety Disorders, 23*, 632-639.

Ghisi, M., Flebus, G. B., Montano, A., Sanavio, E., & Sica C. (2006). *Beck Depression Inventory-Second Edition (BDI-II). Manuale*. Firenze: Organizzazioni Speciali.

Gosling, S. D., & Johnson, J. A. (2010). (Eds.). *Advanced methods for conducting online behavioral research*. Washington, DC: American Psychological Association.

Guo, S., & Fraser, M. W. (2010). *Propensity score analysis. Statistical methods and applications*. Thousand Oaks, CA: Sage.

Ho, D., Imai, K., King, G., & Stuart, E. (2011). MATCHIT: Nonparametric preprocessing for parametric causal inference. Available from http://gking.harvard.edu/node/4355/rbuild_documentation/ (RBuild autobuild user 2011-10-24).

Ho, D., Stuart, E., Imai, K., & King, G. (2011). Package "MatchIt." Available from http://cran.r-project.org/web/packages/MatchIt/index.html (version 2-4-20).

Kim, H. J., Steketee, G., & Frost, R. O. (2001). Hoarding by elderly people. *Health and Social Work, 26*, 176-184.

Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods, 38*, 88-91.

Marsh, H. W. (2007). Application of confirmatory factor analysis and structural equation modeling in sport/exercise psychology. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of sport psychology* (3rd ed., pp. 774-798). Hoboken, NJ: Wiley.

Mataix-Cols, D., Frost, R., Pertusa, A., Clark, L., Saxena, S., Leckman, J. F., Stein, D. J., Matsunaga, H., & Wilhelm, S. (2010). Hoarding disorder: A new diagnosis for *DSM-V*? *Depression and Anxiety*, *27*, 556-572.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30-46.

Millsap, R. E., & Tein, J. Y. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39*(3), 479-515.

Mohammadzadeh, A. (2009). Validation of Saving Inventory-Revised (SI-R): Compulsive

hoarding measure. *Iranian Journal of Psychiatry and Clinical Psychology*, *15*, 33-41.

Mueller, A., Crosby, R., Frost, R. O., Leidel, B., Bleich, S., Glaesmer, H., Osen, B., & de Zwaan, M. (2009). Fragebogen zum zwanghaften Horten (FZH)—Validierung der deutschen Version des Saving Inventory-Revised (SI-R). *Verhaltenstherapie*, *19*, 243-250.

Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Unpublished manuscript, available online at http://www.statmodel.com/papers.shtml.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide. Sixth edition*. Los Angeles, CA: Muthén & Muthén.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory, 3rd edition*. New York: McGraw-Hill.

Pani, R., & Biolcati, R. (1998). *Shopping compulsivo*. Urbino: Edizioni Quattroventi.

Pani, R., & Biolcati, R. (2006). *Le dipendenze senza droghe: Shopping compulsivo, internet e gioco d'azzardo*. Torino: Utet.

R Development Core Team. (2009). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from http://www.R-project.org (ISBN 3-900051-07-0).

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41-55.

Sanavio, E., Bertolotti, G., Michielin, P., Vidotto, G., & Zotti, A. M. (1997). *CBA-2.0. Cognitive behavioural assessment 2.0. Scale primarie*. Firenze: Organizzazioni Speciali.

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*, 475-480.

Schafer, J. L., & Kang, J. D. Y. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods, 13*(4), 279-313.

Sica, C., & Ghisi, M. (2007). The Italian versions of the Beck Anxiety Inventory (BAI) and the Beck Depression Inventory-II (BDI-II): Psychometric properties and discriminant power. In M. A. Lange (Ed.), *Leading-edge psychological tests and testing research* (pp. 27-50). Hauppauge, NY: NOVA Publishers.

Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1983). *The State-Trait Anxiety Inventory of Adults (STAI-T). Manual*. Palo Alto, CA: Mind Garden.

SPSS Inc. (2006). *SPSS user's guide version 13.0. 13th edition*. Chicago, IL: SPSS, Inc.

Steketee, G., Frost, R. O., & Kyrios, M. (2003). Cognitive aspects of compulsive hoarding. *Cognitive Therapy and Research, 27*, 463-479.

Steketee, G., Frost, R. O., Tolin, D., Rasmussen, J., & Brown, T. (2010). Waitlist-controlled trial of cognitive behavior therapy for hoarding disorder. *Depression and Anxiety*, *27*, 476-484.

Suler, J. (2004). The online disinhibition effect. *CyberPsychology and Behavior, 7*, 321-326.

Thordarson, D. S., Radomsky, A. S., Rachman, S., Shafran, R., Sawchuck, C. N., & Hakstian, A. R. (2004). The Vancouver Obsessional Compulsive Inventory (VOCI). *Behaviour Research and Therapy, 42*, 1289-1314.

Tolin, D., Frost, R. O., & Steketee, G. (2010). A brief interview for assessing compulsive hoarding: The Hoarding Rating Scale-Interview (HRSI). *Psychiatry Research*, *178*, 147-152.

Tolin, D. F., Frost, R. O., Steketee, G., Gray, K. D., & Fitch, K. E. (2008). The economic and social burden of compulsive hoarding. *Psychiatry Research, 160*, 200-211.

Tortella-Feliu, M., Fullana, M., Caseras, X., Andión, Ó., Torrubia, R., & Mataix-Cols, D. (2006). Spanish version of the Saving Inventory-Revised (SI-R): Adaptation, psychometric properties, and relationship to personality variables. *Behavior Modification*, *30*, 693-712.

Valence, G., d'Astous, A., & Fortier, L. (1988). Compulsive buying: Concept and measurement. *Journal of Consumer Policy, 11*, 419-433.

Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology, 84*, 608-618.