Measuring Orthorexia Nervosa: Psychometric Limitations of the ORTO-15

María Roncero[1], Juan Ramón Barrada[2] and Conxa Perpiñá[1]

[1] *Universitat de Valencia (Spain)*
[2] *Universidad de Zaragoza (Spain)*

Abstract. Orthorexia nervosa has recently been defined as excessive preoccupation with healthy eating, causing significant nutritional deficiencies and social and personal impairments. The ORTO-15 is the most widely used instrument to evaluate orthorexia nervosa, although previous studies obtained inconsistent results about its psychometric properties, and there are no data on the Spanish version. Thus, the main objective of the present study was to analyze the psychometric properties of the Spanish adaptation of the ORTO-15. In order to cross-validate the results, two independent samples were used (Sample 1: $n = 807$, 74.1% women; Sample 2: $n = 242$, 63.2% women). The results did not support the original recoding and reversal of the items; thus, the original scores were maintained. The analysis of the internal structure showed that the best interpretable solution was unidimensional, and due to low loadings, four items were removed. The internal consistency ($\alpha = .74$) and temporal stability ($r = .92$; $p < .001$) of the final ORTO-11 version were adequate, higher than the 15-item version. The questionnaire showed significant associations with eating psychopathology (EAT-26 and SR-YBC-EDS; range $r = .64 – .29$; $p < .05$). However, this result should be interpreted with caution due to the redundancy observed between the ORTO-15 and the EAT-26. Our results suggest that the psychometric properties of the Spanish version of the ORTO-15 are not adequate. Moreover, the instrument detects people who are on diets, but it is not efficient in detecting the severity of orthorexic behaviors and attitudes. New instruments are needed to continue the study of orthorexia nervosa.

Received 6 June 2016; Revised 1 August 2017; Accepted 3 August 2017

Keywords: eating disorders, orthorexia nervosa, psychometric properties, validation.

The term orthorexia nervosa (ON) was coined by Bratman (1997; Bratman & Knight, 2000) to describe an excessive preoccupation about consuming only biologically pure foods and eating in a healthy way, leading to dietary limitation, malnutrition, and underweight. People with this problem are worried about the quality of food, which for them is more important than its quantity or taste. They avoid having foods considered "impure", unhealthy or contaminated (Moroze, Dunn, Holland, Yager, & Weintraub, 2015; Varga, Dukay-Szabó, Túry, & van Furth, 2013). This preoccupation, when extreme, leads to restricted social lives, as orthoretic people do not eat outside because they do not trust the quality of the food or the way it is cooked. They spend excessive amounts of money and time buying food and planning, preparing, and consuming meals (Koven & Abry, 2015; Varga et al., 2013), and if they do not follow their rigid self-imposed rules, they feel guilty and punish themselves (Moroze et al., 2015; Varga et al., 2013).

Although the interest in ON has increased in recent years and professionals from the field of eating disorders recognize this behavioral pattern in their own practice (Vandereycken, 2011), with descriptions of some clinical cases resembling the definition of ON (Moroze et al., 2015), this concept is still controversial. ON is not recognized as a disorder by the Diagnostic and Statistical Manual of Mental Disorders (5th Ed.; American Psychiatric Association, 2013) or by the International Classification of Diseases (10th Ed.; World Health Organization, 1992). Moreover, there is no consensus about the definition of ON or the possible diagnostic criteria. There have been two attempts to establish diagnostic criteria, by Moroze et al. (2015) and by Dunn and Bratman (2016). Both models conceptualize orthorexia as obsessive preoccupation or focus (which in our opinion would be better conceptualized as an overvalued idea, see Veale, 2002). Moroze's criteria include aspects such as guilt, the excessive amount of money and time consumed, and exclusion criteria (see Moroze et al., 2015) that are not included in Dunn's proposal. Dunn`s criteria include the idea that the accomplishment of the self-imposed eating behavior excessively determines the body image, self-worth, identity and/or satisfaction (Dunn & Bratman, 2016).

Correspondence concerning this article should be addressed to María Roncero. Universidad de Zaragoza - Psicología y Sociología. C/ Ciudad escolar s/n. Teruel (Spain).

E-mail: maria.roncero@uv.es

Some authors point to a symptom overlap between ON and eating disorders, obsessive-compulsive disorder, hypochondriasis, and even psychotic disorders (Brytek-Matera, 2012). Previous studies found significant associations between ON scores and eating disorder symptomatology (Bundros, Clifford, Silliman, & Morris, 2016; Brytek-Matera, Krupa, Poggiogalle, & Donini 2014; Musolino, Warin, Wade, & Gilchrist, 2015; Segura-Garcia et al., 2012, 2015; Stochel et al., 2015) and obsessive-compulsive traits and symptoms (Koven & Senbonmatsu, 2013; Segura-Garcia et al., 2012; 2015). However, further studies are needed to clarify what place ON should take in the spectrum of disordered eating.

To help clarify the research on ON, it is crucial to have valid instruments for its assessment. As far as we know, three instruments have been developed with this goal in mind: Bratman's test (Bratman & Knight, 2000), with no validation studies; the ORTO-15 test (Donini, Marsili, Graziani, Imbriale, & Canella, 2005); and more recently, the Eating Habits Questionnaire (EHQ; Gleaves, Graham, & Ambwani, 2013). The limitation of the EHQ is that it does not include items reflecting the negative emotionality associated with ON, as the sadness, fear, shame, guilt, frustration, self-loathing, and self-punishment, which is experimented by patients when they are forced to eat or faced with unhealthy food. Until now, the ORTO-15 has been translated and validated in several languages, and has been used in most of the studies in the field, as if it were considered the gold-standard in the assessment of ON, independently of its psychometrical properties.

Until now, the ORTO-15 has been translated and validated in several languages, and it has been used in most of the studies in the field because it was considered the gold-standard in the assessment of ON, regardless of its psychometric properties.

The ORTO-15 was developed based on Bratman's Test (Bratman & Knight, 2000). It consists of 15 items with responses on a four-point scale ranging from *Always* to *Never*. Donini et al. (2005) considered that the questionnaire tapped three dimensions (although no factor analysis was provided): Six items addressed a cognitive-rational area (e.g., "When eating, do you pay attention to the calories of the food?"); five items were about the clinical area (e.g., "In the last three months, did the thought of food worry you?"); and four items about the emotional area (e.g., "When you go in a food shop, do you feel confused?"). Four items were reverse scored and, importantly, two items followed a specific scoring method: For Item 1 and Item 13, *Always* = 2, *Often* = 4, *Sometimes* = 3, and *Never* = 1. We will refer to these two items as recoded items. A single total scored is computed by adding the scores on each item.

Lower total scores are indicative of ON, and high scores indicate normal eating behavior.

The ORTO-15 questionnaire was originally created in Italian, and it has been translated and validated in various languages: Turkish (Arusoğlu, Kabakçi, Köksal, & Kutluay Merdol, 2008), Portuguese (Alvarenga et al., 2012), Hungarian (Varga, Thege, Dukay-Szabó, Túry, & van Furth, 2014), Polish (Brytek-Matera et al., 2014; Stochel et al., 2015), and German (Missbach et al., 2015). Five studies have analyzed the internal structure of the ORTO-15, with very different results among them. These studies have used principal components analysis (Alvarenga et al., 2012; Arusoğlu et al., 2008), which is not an appropriate technique to evaluate the internal factor structure of a set of items (e.g., Widaman, 2007), and confirmatory factor analysis (Brytek-Matera et al., 2014; Missbach et al., 2015; Varga et al., 2014), which could be considered a doubtful choice, given that the internal structure was fairly unclear. A summary of the proposed structures can be seen in Table 1. The number of factors retained ranges from one (Varga et al., 2014) to three (Alvarenga et al., 2012; Arusoğlu et al., 2008). In some versions, several items have been dropped due to problematic loadings, leading to shortened versions ranging from nine items (Brytek-Matera et al., 2014; Missbach et al., 2015) to 12 items (Alvarenga et al., 2012). Of the 15 items, nine of them have been dropped in one or more of the validation studies: the four reverse-keyed items, the two items that are not linear, and Items 6, 14, and 15.

The present study had two goals. The first goal, given that there are no data about the ORTO-15 in the Spanish language, was to carry out the adaptation of the ORTO-15 to Spanish and study its psychometric properties. The second, given the inconsistencies in the previous results about the internal structure of the ORTO-15, was to offer additional evidence about its factorial structure. To do so, we studied two independent samples in order to cross-validate our results.

Method

Participants and Procedure

Sample 1

Data were collected through the Internet. The link to the survey was distributed through the e-mail distribution lists of students from the university where the first two authors worked. Participants completed the protocol, which took approximately 20 minutes. After reading the description of the study, participants provided informed consent, where the anonymity of the responses was clearly stated. Participants had to be at least 18 years old to take the survey, and participants who reported having a diagnosis of any mental disorder and/or taking

Table 1. *Internal Structure of the ORTO-15 in Different Validation Studies*

	ITA			TUR			POR			POL		HUN	GER	SPA
	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F1	F1	F1
1) When eating, do you pay attention to the calories of the food? P,RC	×			×								×		×
2) When you go into a food shop do you feel confused? P,R		×		×								×		×
3) In the last 3 months, did the thought of food worry you?		×		×			×				×	×	×	×
4) Are your eating choices conditioned by your worry about your health status?			×	×					×	×		×	×	×
5) Is the taste of food more important than the quality when you evaluate food? P,R	×				×				×	×			×	
6) Are you willing to spend more money to have healthier food? P	×				×			×		×			×	
7) Does the thought about food worry you for more than three hours a day?		×					×	×			×	×	×	×
8) Do you allow yourself any eating transgressions? P,R		×			×						×			
9) Do you think your mood affects your eating behavior? P,R		×		×							×	×		×
10) Do you think that the conviction to eat only healthy food increases self-esteem?			×	×				×		×		×	×	×
11) Do you think that eating healthy food changes your life-style (frequency of eating out, friends, …)?	×			×			×			×		×	×	×
12) Do you think that consuming healthy food may improve your appearance?	×			×				×		×		×	×	×
13) Do you feel guilty when transgressing? P,RC			×	×			×					×		×
14) Do you think that on the market there is also unhealthy food? P	×				×			×			×			
15) At present, are you alone when having meals? P		×			×							×	×	×

ITA = Italian version; TUR = Turkish version; POR = Portuguese version; POL = Polish version; HUN = Hungarian version; GER = German version; SPA = Spanish version; P = problematic item (deleted in a previous version); R = reversed item; RC = recoded item. Shaded cells correspond to items removed from the final version.

psychotropic medication were not included in the study. Individuals did not receive any compensation for their participation in the study.

A total of 837 participants completed the measures. Data from 30 participants were excluded because they reported having a mental disorder. Of the 807 remaining participants, 598 were women (74.1%), and 209 were men (25.9%). The mean age was 23.65 years ($SD = 6.01$, range [18. 66]). Regarding education, 0.1% of the sample reported not having studies or primary studies, 1.1% had secondary studies, 74.8% were university students, and 23.9% had completed university studies.

Sample 2

This sample was recruited using the "snowball" method. Advanced psychology students at the third author's university participated voluntarily and received one course credit for their recruitment efforts. They attended a two-hour seminar where they received training about the purpose of the study and how to present the study and the instruments to prospective participants. Each student individually administered the assessment protocol to three friends and/or family members. Participants were volunteers. Before completing the questionnaires, each participant gave written informed consent. Participants who had current mental health problems and/or were taking psychotropic medication were not included in the study. Two weeks later, each student administered a retest of the ORTO-15 questionnaire to one of the three friends and/or family members. A total of 72 participants provided responses to the ORTO-15 on both test and retest.

A total of 248 participants completed the measures. Data from six participants were excluded because they reported having a mental disorder. Thus, 242 individuals made up this sample, 153 women (63.2%). The mean age

was 24.94 years old (*SD* = 7.07; range [18, 54]). Regarding education, 7% of the sample reported having primary studies, 21.1% had secondary studies, and 71.1% were university students or had completed university studies. The mean BMI was 22.41 (*SD* = 3.30; range [16.94, 35.14]). More than half of the sample were students (62.4%), single (73.6%), and had a self-rated medium socio-economic level (63.3%). The present study has been approved by the corresponding ethics committee.

Instruments

In Sample 1, the following questionnaires were administered:

Sociodemographics

Participants reported their sex, age, and education level.

ORTO-15 (Donini et al., 2005)

This is the questionnaire under study, and it is described in detail in the introduction section. It was translated into Spanish by a group of clinical psychologists with strong English language skills. It was then reviewed by a professional native translator and back-translated into Spanish.

In Sample 2, in addition, these questionnaires were included:

Sociodemographics

As in sample 1, participants in Sample 2 provided information about their sex, age, education level, weight (to the nearest kilogram) and height (to the nearest centimeter). Moreover, they were asked four yes/no questions: "Go on a diet", "practice regular physical exercise", "tobacco consumption ", and "alcohol consumption".

Eating Attitudes Test-26 (EAT-26; Garner, Olmsted, Bohr, & Garfinkel, 1982)

This 26-item questionnaire measures the frequency of the individual's behavior or attitudes about eating disorders. Items are responded to on a 6-point scale ranging from *Always* to *Never*, and responses are recoded to a 4-point scale. Although the internal structure of the EAT-26 is not completely clear (e.g., Koslowsky et al., 1992; Ocker, Lam, Jensen, & Zhang, 2007), three different subscales are usually considered: (a) a Diet dimension score (13 items; e.g., "I am terrified about being overweight"), (b) a Bulimia and Food Preoccupation score (6 items; e.g., "I find myself preoccupied with food"), and (c) an Oral Control score (7 items; e.g., "I avoid eating when I am hungry"). For the present study, the Spanish version was used (Castro, Toro, Salamero, & Guimerá, 1991). Cronbach's alphas for the present sample are shown in Table 4.

Self-Report Yale–Brown Cornell Eating Disorders Scale (SR-YBC-EDS; Bellace et al., 2012)

The Spanish adaptation of the SR-YBC-EDS is composed of two parts. Part 1 assesses symptom frequency in the past month through a 65-item checklist with a 5-point scale ranging from *Never* = 0 to *Always* = 4. Two different scores are usually computed, a score for the preoccupations (21 items; e.g., "Think excessively about the fat content of the food"), and a score for the Rituals (44 items; e.g., "Need to cut each piece of food into a specific size"). Part 2 of the SR-YBC-EDS assesses, on a 5-point scale ranging from *None* = 0 to *Extreme* = 4, the severity of preoccupations and rituals separately, based on four items about time occupied, interference, distress, and degree of control. The Spanish version was used (Perpiñá, Giraldo-O'Meara, Roncero, & Martínez-Gómez, 2015). Cronbach's alphas for the present sample are shown in Table 4.

Analysis

All the analyses, whenever possible, were performed with both samples in order to check whether they cross-validated. We analyzed our data in four steps. First, we studied the scoring scheme for the ORTO-15. As previously noted, all the recoded items and all the reverse-keyed items have been found problematic in at least one validation study. We computed the Pearson correlations between the total score just using the direct items and (a) the items that should be recoded (Items 1 and 3; with and without recoding) and (b) the reverse-keyed items (Items 2, 5, 8, and 9; with non-reversed scores and reversed scores). Evidence favoring the scoring scheme proposed by Donini et al. (2005) will be found if the correlations are higher for recoded and reversed scores, as proposed.

Second, we studied the internal structure of the scale. Considering the inconsistencies in previous studies, we opted for an exploratory factor analysis (EFA). To determine the number of dimensions to be retained, we used parallel analysis, following the recommendation of Garrido, Abad, and Ponsoda (2013), and interpretability of the solutions. Models were analyzed using weighted least squares–mean and variance adjusted (WLSMV). By using this estimator instead of one based on maximum-likelihood, we were able to maintain the categorical nature of the responses and obtain less biased estimates (Finney & DiStefano, 2006; Rhemtulla, Brosseau-Liard, & Savalei, 2012). Goodness-of-fit in all derived models was assessed with the common cut-off values for the fit indices (Hu & Bentler, 1999). Thus, as indications of model fit, we consider whether the comparative fit index (CFI) and Tucker-Lewis index (TLI) have values greater than .95, and whether the root mean square error of approximation (RMSEA) is less than .06.

Third, after defining the internal structure of the scale, and removing items if needed, we evaluated the reliability of the scores by computing Cronbach's alpha and test-retest correlation. Fourth, and finally, we computed correlations between the ORTO scores and BMI, EAT, SR-YBC-EDS scores, and mean differences were evaluated with Cohen's *d*.

EFA was performed with M*plus* 7 (Muthén & Muthén, 1998–2012), without modifying any of its default specifications, so a GEOMIN rotation was applied. Parallel analysis was computed with syntax written in R. For all the other analyses, SPSS 20 was used.

Results

According to Donini et al. (2005), Item 1 and Item 13 are not linear with ON, and so they must be recoded. For both the items and samples, when items are recoded, the correlations with the total score (only with direct items) were not only lower in magnitude, but also negative: Mean correlation before recoding was .50; mean correlation after recoding, –.35 (see Table 2).

For the items that were designed to be reverse-keyed, the correlation with a total score (again, computed with only the direct items) should be negative before reversing and positive after being reversed. Contrary to this expectation, for two out of the four reverse-keyed items, the correlations were negative after being reversed for both samples (mean correlation = –.31). For the remaining two items, the correlations were positive (see Table 2), but very small (mean correlation = .07). Considering these results, in the next steps we kept the original scores of the items, without reversing or recoding.

For both samples, the first two factors in the EFA had eigenvalues that were clearly above their random data counterparts (real versus random eigenvalues –Sample 1/Sample 2–: factor 1 = 5.13/4.05 vs. 1.36/1.70, factor 2 = 1.59/1.94 vs. 1.27/1.54), whereas the third eigenvalue was below its random counterpart (real = 1.17/1.23, random = 1.23/1.42). For Sample 1, model fit could be considered satisfactory [$\chi^2(76) = 275.40$, CFI = .962,

Table 2. *Correlation between the total score derived from the direct items on the ORTO-15 and the recoded and reversed items, according to the scoring scheme*

	Sample 1	Sample 2
Item 1	.48	.42
Item 1 – Recoded	–.33	–.32
Item 13	.61	.48
Item 13 – Recoded	–.48	–.29
Item 2 – Reversed	–.35	–.14
Item 5 – Reversed	.12	.05
Item 8 – Reversed	.05	.07
Item 9 – Reversed	–.45	–.32

TLI = .947, RMSEA = .057]. For Sample 2, model fit was not as good [$\chi^2(76) = 131.35$, CFI = .935, TLI = .911, RMSEA = .055]. The correlation between factors was equal to .32 in Sample 1 and –.06 in Sample 2.

When assessing the interpretability of a two-factor solution, shown in Table 3, we found several problems. Item 14 presented very low loadings in any factor for both samples (maximum loading = .25). For several items, allocation to a single dimension was not possible, as the loadings were very similar in both factors (Item 4 and Item 12 in Sample 1, Item 6 in Sample 2). Finally, the pattern of results was not consistent across samples.

Considering the problems with a two-factor solution, we decided to create a unidimensional version of the test. To do so, we retained a single factor with the scores from Sample 1, and we removed the items with loadings below .30 in this solution. With this cut-off criterion, Items 5, 6, 8, and 14 were discarded (loadings –.10, .29, –.06, .28, respectively). The loadings of the shortened version can be seen in Table 3. For Sample 1, the fit of this model can be considered to be basically satisfactory [$\chi^2(44) = 267.86$, CFI = .954, TLI = .942, RMSEA = .079]. The mean, maximum, and minimum loadings were .87, .31, and .62, respectively.

We checked the stability of the factor solution for the shortened test with Sample 2. Although the model fit approached satisfactory levels [$\chi^2(44) = 84.13$, CFI = .947, TLI = .934, RMSEA = .061], the loadings of two items fell below .30. The loading of Item 2 was .25, and for Item 15, it was .16. Considering that the size of Sample 1 is much larger than Sample 2, to avoid further item elimination, and accepting that this shortened version is not consistent in both samples, we considered that the best solution was a shortened version of the scale with the 11 items selected with Sample 1. We will call this shortened version ORTO-11. For comparability with previous studies, we also computed scores for the ORTO-15 following the indications of the authors.

The reliability of the scores on the ORTO-11, as measured by Cronbach's alpha, was satisfactory for both samples and times (Sample 1 – α = .83; Sample 2, test – α = .74; Sample 2, retest – α = .78). Scores showed a high level of temporal stability, with a Pearson correlation of .92; *p* < .001. When we consider the original ORTO-15, Cronbach's alphas were .20 for Sample 1 and .23 for Sample 2 (test), and .45 (retest); test-retest correlation was equal to .78.

The associations between variables calculated with sample 2 can be seen in Table 4. To simplify the interpretation, we computed scores for the ORTO-11 and ORTO-15, where higher values are indicative of higher ON. The correlation between the ORTO-11 and the ORTO-15 was medium-sized (*r* = .44). When we considered the additional variables measured, in all cases the associations were higher with the ORTO-11 scores

Table 3. *Factor Loadings (Pattern Matrix) for Original Version (ORTO-15) and the Shortened Version (ORTO-11) in the Two Samples*

	Sample 1			Sample 2		
	2FS		1FS	2FS		1FS
	F1	F2	F1	F1	F2	F1
1) When eating, do you pay attention to the calories of the food?	**.51**	.24	**.62**	**.67**	.11	**.62**
2) When you go in a food shop do you feel confused?	**.63**	−.17	**.52**	.24	−.10	.25
3) In the last 3 months, did the thought of food worry you?	**.89**	.01	**.87**	**.78**	−.24	**.85**
4) Are your eating choices conditioned by your worry about your health status?	**.46**	.33	**.61**	**.55**	.08	**.50**
5) Is the taste of food more important than the quality when you evaluate food?	.23	**−.55**	——	−.15	**−.56**	——
6) Are you willing to spend more money to have healthier food?	.04	**.47**	——	**.42**	**.50**	——
7) Does the thought about food worry you for more than three hours a day?	**.87**	−.03	**.84**	**.64**	−.39	**.72**
8) Do you allow yourself any eating transgressions?	.15	**−.34**	——	.01	**−.47**	——
9) Do you think your mood affects your eating behavior?	**.60**	−.02	**.57**	**.38**	−.43	**.46**
10) Do you think that the conviction to eat only healthy food increases self-esteem?	**.39**	**.39**	**.59**	**.49**	.04	**.47**
11) Do you think that eating healthy food changes your life-style (frequency of eating out, friends)?	**.41**	.23	**.52**	**.37**	−.27	**.43**
12) Do you think that consuming healthy food may improve your appearance?	**.38**	**.44**	**.59**	**.58**	.04	**.55**
13) Do you feel guilty when transgressing?	**.71**	.20	**.80**	**.68**	−.00	**.69**
14) Do you think that on the market there is also unhealthy food?	.24	.10	——	.25	−.22	——
15) At present, are you alone when having meals?	**.32**	−.01	**.31**	.11	−.19	.16

Note: 2FS = Two factor solution (for ORTO-15). 1FS = One factor solution (for ORTO-11). F1 = First factor. F2 = Second factor.

Table 4. *Associations Between ORTO-11 and ORTO-15 and Other Variables (Sample 2)*

	ORTO-11	ORTO-15	*M*	*SD*	Alpha	*n*
Correlations						
ORTO-11	——	.44	10.26	4.79	.74	242
ORTO-15	.44	——	22.14	3.78	.23	242
EAT Diet	.58	.26	3.61	5.11	.82	242
EAT Bulimia	.59	.33	0.63	1.33	.53	242
EAT Oral Control	.28	.20	1.63	2.33	.54	242
YBC Preoccupations	.64	.23	17.35	11.80	.91	241
YBC Rituals	.59	.29	14.33	13.60	.90	235
YBC Severity Preoccupations	.35	.06	5.14	3.26	.86	242
YBC Severity Rituals	.29	.10	1.63	2.35	.81	242
BMI	.22	.09	22.41	3.30	——	237
Cohen's *d*						
Sex	−0.29	−0.14	0.37	0.48	——	242
Dieting	1.14	0.49	0.13	0.34	——	242
Exercise	0.18	0.09	0.59	0.49	——	241
Smoking	0.07	0.04	0.24	0.43	——	242
Alcohol	0.29	0.15	0.50	0.50	——	242

Note: ORTO-11 = Reduced version of the ORTO-15 scale; EAT = Eating Attitudes Test; YBC = Self-Report Yale–Brown Cornell Eating Disorders Scale; BMI = body mass index. Underlined values correspond to statistically significant correlations or mean differences ($p < .05$). To facilitate the interpretation of ORTO-11 and ORTO-15 scores, and contrary to Donini et al. (2005), higher values correspond to higher orthorexia nervosa. Sex was coded with a dummy variable, where 0 = *women* and 1 = *men*. Dieting, exercise, smoking, and alcohol were coded with dummy variables, where 0 = *no* and 1 = *yes*.

than with the ORTO-15 scores. Considering only the ORTO-11, this measure presented a high overlap (rs > .50) with the Diet and Bulimia factors of the EAT-26 and with Preoccupations and Rituals of the SR-YBC-EDS. BMI and ORTO-11 scores were positively correlated, $r = .22$. Women and non-consumers of alcohol presented higher scores on the ORTO-11, both Cohen's $d = .29$. Importantly, those on a diet

presented much higher mean ORTO-11 scores than those who were not on a diet, *d* = 1.14. The relationship with smoking or physical exercise was not statistically significant.

Discussion

The ORTO-15 is the most widely used instrument to measure ON across the studies. Results about its internal structure have been inconsistent across different validation studies in several languages. Therefore, the main objective of the present study was to analyze the psychometric properties of this instrument. In doing so, we counted on two independent samples from a non-clinical population. Moreover, no data about the Spanish version of this questionnaire have been published.

First, the scoring scheme of the questionnaire was studied. We found no evidence supporting the advisability of recoding some items, and some items that were supposed to be reversed offered higher correlations with the direct items when non-reversed. This original scoring scheme has also been found to be problematic in previous studies (Alvarenga et al., 2012).

Regarding the internal structure of the questionnaire, in a first step, a two-factor solution was found. However, due to interpretability problems (i.e., several items had similar loadings in both factors, and the patterns of results were inconsistent in the two samples), we opted for a unidimensional solution. To do so, four items had to be removed, leading to the ORTO-11. This version of the ORTO questionnaire corresponds to the one found in the Hungarian version. Moreover, as Table 1 shows, Items 8 and 14 have repeatedly been deleted in the Polish, German, and Hungarian versions.

When the stability of the one-factor solution of the ORTO-11 was tested in Sample 2, the model fit was satisfactory, but two items showed loadings lower than .30: items 2 and 15. These items have been found to be problematic in other validations (see Table 1). Item 2 was deleted in the Portuguese, Polish, and German versions, whereas Item 15 was deleted in the Polish version.

The internal consistency with the 11 items was adequate (α = .74). This result agrees with the Hungarian version (α = .82), which shares the same structure. However, in the Portuguese, Turkish, and German versions, with other item structures, the internal consistency was not adequate (α < .70). The Cronbach's alpha of our proposed version of the ORTO is much higher than the Cronbach's alpha of the full version with item reversal and recoding. Moreover, our shortened version presented higher temporal stability.

The ORTO-11 was associated with external variables to a larger degree than the original ORTO-15. We consistently found that higher eating psychopathology scores were associated with higher scores on the ORTO-11, mainly Diet and Bulimia factors of the EAT-26 and Preoccupations and Rituals factors of the SR-YBC-EDS. This pattern of results suggesting the association with eating psychopathology agrees with previous studies (Barnes & Caltabiano, 2016; Brytek-Matera et al., 2014; Bundros et al., 2016; Sanlier, Yassibas, Bilici, Sahin, & Celik, 2016).

However, this result about the association between ON and eating disorders should be interpreted with caution, due to the redundancy observed between the ORTO-15 and EAT-26. For instance, the Item 1 wording – "When eating, do you pay attention to the calories of the food?"– presents a high overlap with the wording of Item 6 on the EAT-26 ("I am aware of the calorie content of foods that I eat"), but this item on the EAT-26 is supposed to measure dieting, not ON. There is also a similar overlap between Item 7 on the ORTO-15 –"Does the thought of food worry you for more than three hours a day?"– and items 3 and 21 on the EAT-26 –"Find myself preoccupied with food" and "Give too much time and thought to food"–.

The ORTO-11 was associated with a measure that evaluates preoccupations and rituals related to eating disorders, the SR-YBC-EDS. Results showed that the severity of preoccupations and rituals was associated with orthorexia in a positive way, i.e., the greater the severity of the preoccupations and rituals, the greater the severity of ON. However, the association was lower when compared to the indices for the association between ORTO-11 and EAT-26. The association between SR-YBC-EDS and ORTO has been observed in previous studies (Segura-García et al., 2012; Segura-García et al., 2015) and agrees with the suggestion that ON is associated with some obsessive-compulsive traits (Bundros et al., 2016; Brytek-Matera, 2012; Koven & Abry, 2015; Koven & Senbonmatsu, 2013; Varga et al., 2013).

Integrating our results with those from previous studies, mainly those from Missbach et al. (2015), we consider that the use of the ORTO-15, and the shortened versions derived from it, should be discontinued. The performance of this instrument has been inconsistent across samples, with almost as many versions as studies. The method for computing a total score is not clear, as the items that should be reversed or recoded are found to be problematic or offer better performance without incorporating the changes suggested by Donini et al. (2005). The content validity of the instrument is doubtful. The high correlations with EAT-26 and SR-YBC-EDS scores, the theoretically unexpected correlation between the ORTO-11 and BMI, and the very large mean difference between dieters and non-dieters, can be easily interpreted if we consider that, instead of assessing ON, the ORTO-11 is mainly measuring dieting or restrictive eating. The satisfactory reliability of the ORTO-11 scores indicates that this scale measures something with a small error, but

the other available data suggest that we are not measuring what we are supposed to be measuring, ON.

The present study has some limitations. Data are based on self-reports, and both samples were convenience samples mainly composed of university students. The questions about going on a diet, doing exercise, or alcohol and tobacco consumption were yes/no questions, and in future studies this should be specified to better understand the ON pattern in non-clinical populations. Cronbach's alpha for the Bulimia and Oral Control scales were rather low, probably due to a floor effect. And finally, our Spanish version of ORTO-15 was translated from the English version, and not directly from the Italian version.

This study represents an addition effort to evaluate the validity of the ORTO-15. We have done so with two independent samples and starting from the basis: the scoring scheme, internal structure with an exploratory analysis, and relation with other variables. The validity of the instrument has been questioned by other authors (see Dunn & Bratman, 2016; Koven & Abry, 2015; Moroze et al., 2015; Varga et al., 2014), and our results agree with these previous studies, suggesting that the psychometric properties of the ORTO-15 are not adequate. Our results indicate that the ORTO-11 scores detect people who are on diet, but this instrument is not efficient in detecting the severity of orthorexic behaviors and attitudes. With all of this in mind, greater attention is needed on other measures of ON (e.g., the EHQ) or on the development of new instruments. Furthermore, much of what is supposed to be known about ON should be reconsidered, as the ORTO, used in most of the studies as if it were considered the gold-standard, is not very golden.

References

Alvarenga M. S., Martins M. C. T., Sato K. S. C. J., Vargas S. V. A., Philippi S. T., & Scagliusi F. B. (2012). Orthorexia nervosa behavior in a sample of Brazilian dietitians. *Eating and Weight Disorders*, 17, 29–35. https://doi.org/10.1007/BF03325325

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th Ed.). Washington, DC: Author.

Arusoğlu G., Kabakçi E., Köksal G., & Kutluay Merdol T. (2008). Orthorexia nervosa and adaptation of ORTO-11 into Turkish. *Turkish Journal of Psychiatry*, 19, 1–9.

Barnes M. A., & Caltabiano M. L. (2017). The interrelationship between orthorexia nervosa, perfectionism, body image and attachment style. *Eating and Weight Disorders*, 22, 177–184. https://doi.org/10.1007/s40519-016-0280-x

Bellace D. L., Tesser R., Berthod S., Wisotzke K., Crosby R. D., Crow S. J., … Halmi K. A. (2012). The Yale-Brown-Cornell eating disorders scale self-report questionnaire: A new, efficient tool for clinicians and researchers. *International Journal of Eating Disorders*, 45, 856–860. https://doi.org/10.1002/eat.22023

Bratman S. (1997). Health food junkie. *Yoga Journal*, 136, 42–50.

Bratman S., & Knight D. (2000). *Health food junkies*. New York, NY: Broadway Books.

Brytek-Matera A. (2012). Orthorexia nervosa—an eating disorder, obsessive-compulsive disorder or disturbed eating habit? *Archives Psychiatry Psychotherapy*, 14, 44–60.

Brytek-Matera A., Krupa M., Poggiogalle E., & Donini L. M. (2014). Adaptation of the ORTHO-15 test to Polish women and men. *Eating and Weight Disorders*, 19, 69–76. https://doi.org/10.1007/s40519-014-0100-0

Bundros J., Clifford D., Silliman K., & Morris M. N. (2016). Prevalence of Orthorexia Nervosa among college students based on Bratman's test and associated tendencies. *Appetite*, 101, 86–94. https://doi.org/10.1016/j.appet.2016.02.144

Castro J., Toro J., Salamero M., & Guimerá E. (1991). The eating attitudes test: Validation of the Spanish version. *Evaluación Psicológica*, 7, 175–189.

Donini L. M., Marsili D., Graziani M. P., Imbriale M., & Cannella C. (2005). Orthorexia nervosa: A preliminary study with a proposal for diagnosis and an attempt to measure the dimension of the phenomenon. *Eating and Weight Disorders*, 9, 151–157. https://doi.org/10.1007/BF03325060

Dunn T. M., & Bratman S. (2016). On orthorexia nervosa: A review of the literature and proposed diagnostic criteria. *Eating Behaviors*, 21, 11–17. https://doi.org/10.1016/j.eatbeh.2015.12.006

Finney S. J., & DiStefano C. (2006). *Non-normal and categorical data in structural equation modeling*. In G. R. Hancock & R. O. Mueller (Eds.), Structural equation modeling: A second course. Greenwich, CT: Information Age Publishing.

Garner D. M., Olmsted M. P., Bohr Y., & Garfinkel P. E. (1982). The eating attitudes test: Psychometric features and clinical correlates. *Psycholological Medicine*, 12, 871–878. https://doi.org/10.1017/S0033291700049163

Garrido L. E., Abad F. J., & Ponsoda V. (2013). A new look at Horn's parallel analysis with ordinal variables. *Psychological Methods*, 18, 454–474.

Gleaves D. H., Graham E. C., & Ambwani S. (2013). Measuring "orthorexia": Development of the Eating Habits Questionnaire. *International Journal of Educational and Psychological Assessment*, 12, 1–18.

Hu L. T., & Bentler P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. https://doi.org/10.1080/10705519909540118

Koslowsky M., Scheinberg Z., Bleich A., Mark M., Apter A., Danon Y., & Solomon Z. (1992). The factor structure and criterion validity of the short form of the Eating Attitudes Test. *Journal of Personality Assessment*, 58, 27–35. https://doi.org/10.1207/s15327752jpa5801_3

Koven N. S., & Abry A. W. (2015). The clinical basis of orthorexia nervosa: Emerging perspectives. *Journal of Neuropsychiatric Disease and Treatment*, 18, 385–394. https://doi.org/10.2147/NDT.S61665

Koven N. S., & Senbonmatsu R. (2013). A neuropsychological evaluation of orthorexia nervosa. *Open Journal of Psychiatry*, 3, 214–222. https://doi.org/10.4236/ojpsych.2013.32019

Missbach B., Hinterbuchinger B., Dreiseitl V., Zellhofer S., Kurz C., & König J. (2015). When eating right, is measured

wrong! A validation and critical examination of the ORTO-15 questionnaire in German. *PLoS ONE, 10,* e0135772. https://doi.org/10.1371/journal.pone.0135772

Moroze R. M., Dunn T. M., Holland J. C., Yager J., & Weintraub P. (2015). Microthinking about micronutrients: A case of transition from obsessions about healthy eating to near-fatal ''orthorexia nervosa'' and proposed diagnostic criteria. *Psychosomatics, 56,* 397–403. https://doi.org/10.1016/j.psym.2014.03.003

Musolino C., Warin M., Wade T., & Gilchrist P. (2015). "Healthy anorexia": The complexity of care in disordered eating. *Social Science & Medicine, 139,* 18–25. https://doi.org/10.1016/j.socscimed.2015.06.030

Muthén L. K., & Muthén B. O. (1998–2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

Ocker L. B., Lam E. T. C., Jensen B. E., & Zhang J. J. (2007). Psychometric properties of the Eating Attitudes Test. *Measurement in Physical Education and Exercise Science, 11,* 25–48. https://doi.org/10.1080/10913670709337010

Perpiñá C., Giraldo-O'Meara M., Roncero M., & Martínez-Gómez M. (2015). Confirmatory factor analysis and psychometric properties of the Yale–Brown–Cornell Eating Disorders Scale Self-Report version (SR-YBC-EDS) in Spanish clinical and non-clinical samples. *Eating Behaviors, 17,* 6–9. https://doi.org/10.1016/j.eatbeh.2014.12.005

Rhemtulla M., Brosseau-Liard P. É., & Savalei V. (2012). When can categorical variables be treated as continuous?. A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17,* 354–373. https://doi.org/10.1037/a0029315

Sanlier N., Yassibas E., Bilici S., Sahin G., & Celik B. (2016). Does the rise in eating disorders lead to increasing risk of orthorexia nervosa? Correlations with gender, education, and body mass index. *Ecology Food Nutrition, 55,* 266–278. https://doi.org/10.1080/03670244.2016.1150276

Segura-Garcia C., Papaianni M. C., Caglioti F., Procopio L., Nistico C. G., Bombardiere L., … Capranica L. (2012). Orthorexia nervosa: A frequent eating disordered behavior in athletes. *Eating and Weight Disorders, 17,* 226–233.

Segura-Garcia C., Ramacciotti C., Rania M., Aloi M., Caroleo M., Bruni A., … De Fazio P. (2015). The prevalence of orthorexia nervosa among eating disorder patients after treatment. *Eating and Weight Disorders, 20,* 161–166. https://doi.org/10.1007/s40519-014-0171-y

Stochel M., Janas-Kozik M., Zejda J. E., Hyrnik J., Jelonek I., & Siwiec A. (2015). Validation of ORTO-15 Questionnaire in the group of urban youth aged 15–21. *Psychiatria Polska, 49,* 119–134. https://doi.org/10.12740/PP/25962

Vandereycken W. (2011). Media hype, diagnostic fad or genuine disorder? Professionals' opinions about night eating syndrome, orthorexia, muscle dysmorphia, and emetophobia. *Eating Disorders, 19,* 145–155. https://doi.org/10.1080/10640266.2011.551634

Varga M., Dukay-Szabó S., Túry F., & van Furth E. F. (2013). Evidence and gaps in the literatura on orthorexia nervosa. *Eating and Weight Disorders, 18,* 103–111. https://doi.org/10.1007/s40519-013-0026-y

Varga M., Thege B. K., Dukay-Szabó S., Túry F., & van Furth E. F. (2014). When eating healthy is not healthy: Orthorexia nervosa and its measurement with the ORTO-15 in Hungary. *BMC Psychiatry, 14*(1), 1. https://doi.org/10.1186/1471-244X-14-59

Veale D. (2002). Over-valued ideas: A conceptual analysis. *Behaviour Research and Therapy, 40,* 383–400. https://doi.org/10.1016/S0005-7967(01)00016-X

Widaman K. (2007). Common factors versus components: Principals and principles, errors and misconceptions. In R. Cudeck & R. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 177–203). Mahwah, NJ: Erlbaum.

World Health Organization. (1992). *International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10)*. Geneva, Switzerland: WHO.