THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

The psychometric properties of Orthorexia Nervosa assessment scales

Citation for published version:
Opitz, M-C, Newman, E, Alvarado Vázquez Mellado, S, Robertson, M & Sharpe, H 2020, 'The psychometric properties of Orthorexia Nervosa assessment scales: A systematic review and reliability generalization', *Appetite*, vol. 155, 10479. https://doi.org/10.1016/j.appet.2020.104797

Digital Object Identifier (DOI):
10.1016/j.appet.2020.104797

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Appetite

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 16. Aug. 2022

The Psychometric Properties of Orthorexia Nervosa Assessment Scales: A Systematic Review and Reliability Generalization.

Marie-Christine Opitz[a*], Emily Newman[b], Ana Sofía Alvarado Vázquez Mellado[c], MacKenzie Robertson[d], Helen Sharpe[e]

[a] Department of Clinical Psychology, School of Health in Social Science, The University of Edinburgh, Medical School (Doorway 6), Teviot Place, EH8 9AG, Scotland, marie-christine.opitz@ed.ac.uk
[b] Department of Clinical Psychology, School of Health in Social Science, The University of Edinburgh, 18 Buccleuch Place, EH8 9LN, Scotland, emily.newman@ed.ac.uk
[c] The Scottish Collaboration for Public Health Research (SCPHRP), The University of Edinburgh, 5 Forrest Hill EH1 2QL, Scotland, sofia.alvarado@ed.ac.uk
[d] Department of Clinical Psychology, School of Health in Social Science, The University of Edinburgh, Medical School (Doorway 6), Teviot Place, EH8 9AG, Scotland, m.d.a.robertson@sms.ed.ac.uk
[e] Department of Clinical Psychology, School of Health in Social Science, The University of Edinburgh, Room 2.8, Doorway 6, Medical Quad, Teviot Place, EH8 9AG, Scotland, helen.sharpe@ed.ac.uk

Abstract

The main objective of this systematic review was to provide a comprehensive overview of the psychometric properties of all available Orthorexia Nervosa (ON) assessment tools, in order to evaluate their scope of application for research and practice. Ten databases were searched for studies quantitatively assessing ON. The psychometric properties were evaluated according to specified quality criteria, focusing on the reliability, structural validity and construct validity of the scales. A meta-analytic approach was used to summarize eligible Cronbach's alpha coefficients between studies. Sixty-eight unique studies fulfilled the inclusion criteria for this systematic review. Ten discrete ON scales were identified. Half of the included studies exclusively utilized a version of the ORTO-15. The evaluation of all available ON measures raise issues regarding ON's dimensionality and conceptualization. Most of the identified scales require further validation. Based on the reported psychometric properties it is advised to re-evaluate existing tools and to focus on establishing consensus regarding the conceptualization of ON to establish a measure with sound psychometric properties.

* Correspondence to Marie-Christine Opitz, Department of Clinical Psychology, School of Health in Social Science, The University of Edinburgh, Medical School (Doorway 6), Teviot Place, EH8 9AG, Scotland
Email: marie-christine.opitz@ed.ac.uk

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

Keywords: Orthorexia Nervosa, Psychometrics, Systematic Review, Reliability Generalization

1 Introduction

1.1 Rationale

Orthorexia Nervosa (ON) has been defined as a pathological obsession, fixation or preoccupation with healthy food (e.g. Andreas et al., 2018; Barrada & Roncero, 2018; Barthels, Meyer & Pietrowsky, 2015b; Bauer et al., 2019; Brytek-Matera et al., 2014; Chard et al., 2018; Glen & Gleaves, 2018; Haddad et al., 2019; He et al., 2019; Rogoza, 2019), a new eating disorder (Donini et al., 2005) or "[…] an otherwise healthy behavior […] taken to extremes […]" (Gleaves et al., 2013, p. 1). However, ON has so far not been recognized by the Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM-5; American Psychiatric Association, 2013) or the international statistical classification of diseases and related health problems (ICD-11; World Health Organization, 2018).

Despite the lack of universally accepted diagnostic criteria, ON has been studied increasingly in the last two decades, which resulted in the publication of four classification approaches defining possible diagnostic criteria (Setnick, 2013; Moroze et al., 2015; Barthels, Meyer & Pietrowsky, 2015b; Dunn & Bratman, 2016). All these approaches refer to an obsessional or pathological preoccupation with healthy nutrition, emotional consequences for transgressing self-imposed dietary rules and psychosocial impairments due to ON (Cena et al., 2018). However, in direct comparison the four approaches differ regarding individual criteria, including conceptual contradictions. For example, while Setnick's (2013) approach includes 'phobic avoidances' and an interrelation between a restrictive diet and an ostensible

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

medical condition, Moroze et al. (2015) suggest that ON should only be diagnosed if a specialized diet is not related to diagnosed food allergies or medical conditions calling for this diet. Furthermore, Barthels, Meyer and Pietrowsky's (2015b) criteria specify that an *intended* weight loss and underweight may be present, while Dunn and Bratman (2016) include the absence of a desire to lose weight as essential for the diagnosis of ON.

Discrepancies in the conceptualization of ON impact the validity of ON measures. Existing ON scales, or the lack of a standardized measure, have been criticized by multiple reviews (e.g. Valente, Syurina & Donini, 2019; Costa, Hardan-Khalil & Gibbs, 2017; Missbach, Dunn & König, 2017; Missbach et al., 2015; Koven & Abry, 2015). However, no review so far has systematically evaluated the psychometric properties of all the available measures. In order for ON research to move forward, it needs to be clear how existing measures perform in comparison to one another.

1.2 Objectives

This systematic review has two main aims. First, to identify all quantitative measures assessing ON, and second to evaluate these measures' psychometric properties.

2 Methods

2.1 Protocol and Registration

The review protocol was registered on International prospective register of systematic reviews (PROSPERO, registration number CRD42019131090) in April 2019, in order to ensure its transparency and quality standards (Booth, 2012; Sideri, Papageorgiou, & Eliades, 2018). The protocol was updated on 17th December to include a reliability generalization (RG) analysis. The RG analysis was added to the

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

protocol as it provides additional information for the analysis of psychometric properties.

2.2 Eligibility Criteria

For this systematic review, all studies assessing ON with a quantitative measure were included, as long as they reported at least one of the psychometric properties specified as part of this review (Internal Consistency, Re-test Reliability, Structural Validity and Construct Validity as determined in relation to established measures). ON had to be assessed with one of the following methods: the application of a questionnaire, inventory, single-question, scale or subscale. Published literature, grey literature (OpenGrey) and master's as well as doctorial theses (ProQuest Dissertations & Theses Global) were eligible for screening. No limit was set regarding the publication date and all articles written in English, German, French, Dutch and Spanish were included. These languages were chosen based on the language proficiency of the first author (MCO). Any other languages were excluded (n = 12) as no translation software was used in order to avoid misinterpretations. Excluded were reviews (n = 12) and studies reporting the same results in more than one publication (n = 8), to avoid multiple publication bias.

2.3 Information Sources

EMBASE, PubMed/MEDLINE, Scopus, PsycInfo, Web of Science, ProQuest Dissertations & Theses Global, ASSIA, CINAHL, OpenGrey and ETHOS were searched for titles and abstracts, once in April 2019 and once in August 2019, in order to update the literature search.

2.4 Search Strategy

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

The search terms were based on a pilot screening of titles and abstracts, which identified relevant questionnaires and spellings. These were specified as “Orthorexi*” (accounting for English and German spelling), “Ortore*a” (accounting for Italian, Spanish and Turkish spelling), “Ortorexi” (Swedish), “obsessive healthy eating”, “ORTO-15”, “ORTO-11”, “ORTO-9”, “EHQ”, “Eating Habits Questionnaire” and “Bratman Test”. References and citations were used to identify additional relevant articles. Once duplicates were deleted, missing data was identified and authors were contacted, up to two times, via email if a contact address was provided on the paper or via ResearchGate.

2.5 Study Selection

All studies fulfilling the eligibility criteria were included in the systematic review. For the meta-analytic RG, all studies reporting a Cronbach’s alpha value for the total scale of ON measures were included if at least two values were reported by a minimum of two independent studies (Higgins et al., 2019).

2.6 Data Collection Process

The online tool Covidence was used to coordinate the screening process between the first (MCO) and second (ASAVM) reviewer. Both reviewers screened titles and abstracts on Covidence after extracting them from the search databases. The included texts were assessed in their full-text version by the first author. The second reviewer independently assessed a random sample (20% of the results). Conflicting decisions were discussed in order to reach agreement. A third independent reviewer (MR) screened a different random sample (20% of the final sample) of the full texts included. Psychometric properties were documented in a table to identify the outcomes relevant for this review.

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

2.7 Outcomes

The studies included were evaluated based on an adapted version of the Quality Criteria for Measurement Properties of Health Status Questionnaires formulated by Terwee et al. (2007) and the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Risk of Bias Checklist (Prinsen et al., 2018; Mokkink et al., 2018; Terwee et al., 2018). Every measure identified was outlined in alphabetical order (Table 1). The evaluation of measurement properties focused on the content validity (conceptual framework, measurement aim, target population and item selection strategy), internal consistency, re-test reliability, structural validity (dimensionality) and construct validity (associations with any other measure) of the individual scales. Internal consistency values were interpreted based on accepted standards (DeVellis, 2003; Nunnally, 1967 in Tavakol & Dennick, 2011).

2.8 Data Synthesis

The results of the systematic evaluation of measurement properties were summarized by measure (the evaluation process is illustrated in the supplementary materials). A summary table was created to compare the findings of all included studies (Table 2). In order to estimate the overall reliability of tests scores for the same measure between studies, an RG was conducted for the Cronbach's alpha values reported. In the present study, a meta-analysis of available reliability coefficients was conducted following Rodriguez and Maeda's (2006) discussion on the "Meta-Analysis of Coefficient Alpha". Based on this study, Cronbach's alpha was transformed using the transformation $T_i = (1 - r_{\propto i})^{\frac{1}{3}}$ by Hakstian and Whalen (1976) (with r_α being the sample coefficient alpha and T_i being the transformed alpha value). This approach was chosen due to the results of Rodriguez and Maeda's comparison of common RG approaches, which showed that T_i exhibited the smallest

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

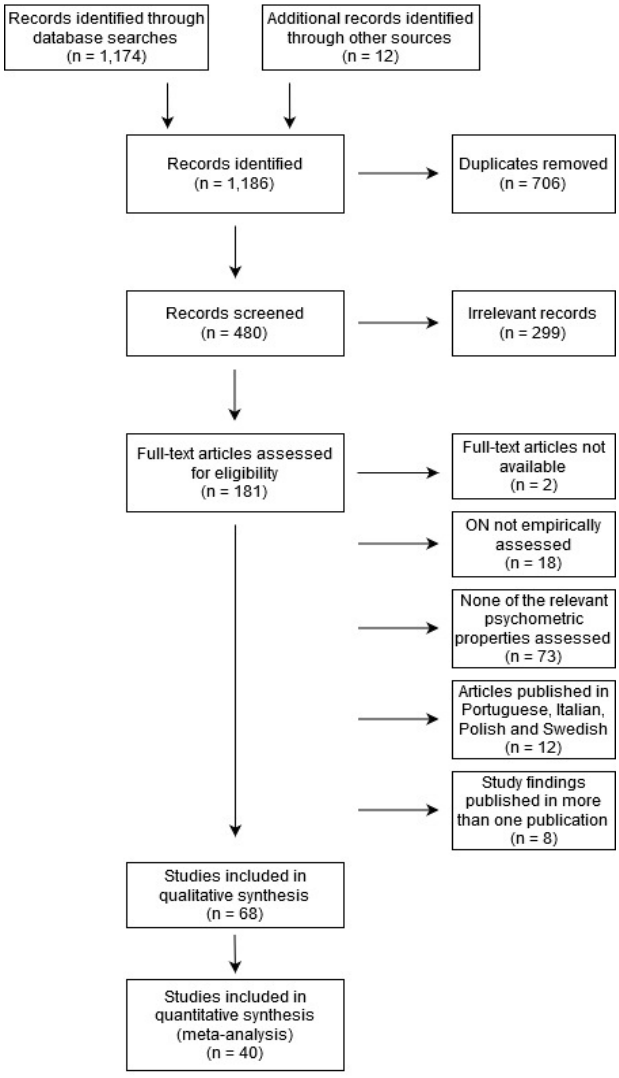
standard error in comparison to an unweighted mean alpha and a variance-adjusted alpha coefficient. R_{Studio} was used to calculate the effect sizes using a random-effects model with the R-code "AHW" for transformed alpha values, as outlined in https://cran.r-project.org/web/packages/metafor/metafor.pdf.

3 Results

3.1 Study Selection

In total, 1,174 studies were identified through the database searches. Eight further studies were identified by cross-checking the references of included articles and four unpublished studies were provided by respective authors. Seven-hundred-and-six duplicates were removed. Consequently, the first (MCO) and second (ASAVM) reviewer screened 480 study titles and abstracts. This process identified 299 studies as irrelevant for this review, as they did not meet the inclusion criteria (e.g., reports, letters to the editor, case studies etc.). As a result, 181 full texts were assessed for eligibility. The data extraction was consequently based on 68 unique studies for the narrative review and 40 studies for the RG analysis. A flowchart (Figure 1) depicts the details of the search process.

Figure 1 Flowchart illustrating the study selection process



© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

Most of the studies included in this review used one version of the ORTO (50%), 11.8% used the EHQ, 10.3% a version of the DOS, 4.4% the BOT and 13.2% a combination of measures.

3.2 Psychometric properties

Information on psychometric properties were available for 10 distinct ON measures (Table 1), which varied in their dissemination and utilization. A table illustrating the psychometric evaluation of all ON measures can be found in the supplementary material. Table 2 summarizes the findings per measure.

3.2.1 Body-Image Screening Questionnaire (BISQ, k=2)

Five items of the BISQ assess ON tendencies. The internal consistency for this subscale was identified as insufficient (Cronbach's alpha=.59) and acceptable (α=.77) in two different studies, which were conducted in two different countries.

3.2.2 Burda-Orthorexia Risk Assessment (B-ORA, k=1)

The B-ORA was developed as part of a doctoral thesis and has not been evaluated beyond this original study. The internal consistency appeared to be high (Cronbach's alpha=.97) and a single factor was identified after adjusting for error terms. A moderate positive correlation was found with disordered eating attitudes.

3.2.3 Bratman Orthorexia Test (BOT, k=5)

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

The original questions of the BOT were developed as a personal risk-assessment for people overly concerned with a healthy diet. As measurement theory was not considered for the construction of the BOT, the scale's three different language adaptations utilize differing score interpretations to assess ON. The reported score values for internal consistency range from α=.67 up to α=.79 for Cronbach's alpha (.60 for the Kuder-Richardson Formula 20). This causes concerns for the scale's reliability, as half of the studies reported a value below .7. The RG analysis revealed an acceptable population alpha of α_p=.73, which is based on three alpha values reported for the 10-item version of the BOT. Test-retest reliability has not been assessed so far. The whole scale is moderately to strongly and positively correlated with a measure for disordered eating behavior and eating disorder related eating patterns. One study identified 5 eating disorder specific and 4 ON-specific items for the 9-item version of the scale. The ON-specific items were only weakly correlated with the same measure of disordered eating.

3.2.4 Düsseldorf Orthorexia Scale (DOS, k=10)

The population alpha for the DOS was assessed using 11 Cronbach's alpha values for the 10-item scale, which revealed a coefficient of α_p=.85 (Cronbach's alpha ranging from α=.69 to α=.93). The internal consistency was re-tested within one sample in three-months intervals, which identified consistently acceptable values between α=.79 and α=.84. The test re-test reliability ranged from r=.67 (first 3 months interval) to r=.77 (4 weeks) and r=.79 (second 3 months interval) for the total scale, indicating variable repeatability. The dimensionality of the DOS remained questionable: a single-factor, 3-factorial and 5-factorial structure were found, but poor model fit indices were common in the studies evaluating the DOS. The scale has been shown to be highly positively related to "Drive for Thinness" (Eating

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

Disorder Inventory: EDI, EDI-2), small to moderately to “Body Dissatisfaction” (EDI-2) and variably to “Bulimia” (EDI-2). A high positive correlation was further found with inflexible eating. Hypochondriacal worries and beliefs, as well as “Cognitive Restraint” (Three-Factor-Eating-Questionnaire-Revised: TFEQ-R18) were positively, but only weakly related to the DOS.

3.2.5 Eating Habits Questionnaire (EHQ, k=10)

The total EHQ shows continuously good Cronbach’s alpha values for the total test scores (α=.86 up to α=.94), with a population alpha α_p=.85 for six reported alpha values. The individual subscales slightly differ in their values, which is partly due to the different labels and items authors assign to the subscales (Cronbach’s alpha ranging from α=.70 for “Feelings” up to α=.92 for “Problems”). Test re-test reliability (after 2-4 weeks) had only been assessed by one study, which found values of r>.70 for all subscales. Even though a 3-factorial structure was identified by three out of four studies, the item-scale allocation remains inconclusive, based on the reported factor loadings. The total EHQ exhibits small positive correlations with anxiety, depression, perfectionism and narcissistic personality traits. The relationship with measures for disordered eating are less clear, even though consistently positive, ranging from small and moderate to high, with some inconsistencies among studies for the Bulimia-Test Revised (BULIT-R) and the Eating Attitudes Test (EAT-26). The “Problems” scale showed higher correlations with disordered eating (EAT-26, BULIT-R subscales “Body Image/Weight Loss”, “Vomiting/Laxatives”) than the other subscales, a moderate correlation with perfectionism and a weak to moderate correlation with depression. All subscales showed only weak correlations with narcissistic personality traits.

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

246 **3.2.6 Eating Habits Questionnaire – Revised (EHQ-R, k=2)**

247 The revised EHQ, EHQ-R, is in the early stages of its evaluation. The two studies
248 which assessed its internal consistency reported good Cronbach's alpha values for
249 the total scale and four of the five subscales, with an acceptable Cronbach's alpha
250 value for the subscale "Time Impairment". The suggested 5-factorial structure needs
251 further validation. The same applies for the scale's construct validity. The authors of
252 the tool found large correlations with disordered eating behaviors for the total scale
253 and varying correlation coefficients for the EHQ-R subscales.

254

255

256 **3.2.7 Orthorexia Nervosa Scale (ONS, k=1)**

257 The ten subscales of the ONS showed almost entirely acceptable to excellent
258 internal consistency values (Cronbach's alpha=.70 to α=.92), except for the subscale
259 "Fasting" (α=.69). The measure's author identified the 10-factorial structure with an
260 underlying second-order factor, suggesting that the total ONS scale is measuring
261 one underlying construct. A small positive correlation with food addiction was found
262 for the total scale.

263 **3.2.8 ORTO-15 (k=37)**

264 The most commonly used questionnaire for the assessment of ON was the ORTO-
265 15. So far, the scale has been adapted for seven languages. Nine different ORTO
266 versions were developed by excluding individual items with the aim of improving
267 either the scale's face validity, internal consistency, model fit or factorial
268 interpretability. The reported internal consistency values (Cronbach's alpha) for test

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

scores range from as low as α=.14 (ORTO-15) up to α=.86 (ORTO-11). Based on the RG with 24 reported Cronbach's alpha values, the population alpha for the 15-item version following the original scoring instructions was identified as α_p=.62. Even though recommendations for the satisfactory level of Cronbach's alpha values differ, an alpha value below α=.70 is consistently seen as questionable (DeVellis, 2003; Bland & Altman, 1997; Nunnally, 1967 in Tavakol & Dennick, 2011). The test re-test reliability for the individual items of the ORTO-15 had only been assessed by two studies using Cohen's kappa, suggesting mixed results. However, the kappa coefficient was designed to measure observer agreement (Landis & Koch, 1977) rather than score repeatability. Using Cohen's kappa for the estimation of test re-test reliability violates the assumption of independent raters (Cohen, 1960) and is therefore not conclusive.

The ORTO-15 was designed to measure three dimensions: cognitive-rational, clinical and emotional aspects of ON. However, the factorial structure of the questionnaire remains uncertain. A single-factor, 2-factorial and 3-factorial structure have been proposed for different item-lengths of the scale. Only one study had evaluated the original 15-item version regarding its dimensionality, which identified a 3-factorial structure for the scale. All other studies evaluating the ORTO-15 reduced the item-length in order to improve its model fit.

Regarding the ORTO's construct validity, the only pattern identifiable was its consistent negative correlation with established tools measuring disordered eating. Higher ORTO-scores indicate less ON tendencies. This interpretation is not consistent across all included studies, making the interpretation of associations difficult. The values cover low, moderate and even large correlations, depending on the ORTO version utilized. Greater ON tendencies were weakly to moderately

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

associated with higher symptoms of depression and OCD (for shorter item-versions of the ORTO).

3.2.9 Scale to Measure Orthorexia in Puerto Rican Men and Women (k=1)

The internal consistency for the Puerto Rican ON scale has only been assessed by the measure's authors. The Cronbach's alpha values range from questionable (α=.66 for "Lifestyle") to good (α=.84 for "Food Intake"). The total scale had a good internal consistency (α=.87). The authors identified a 5-factorial structure. No information was available on the measure's construct validity.

3.2.10 Teruel Orthorexia Scale (TOS, k=2)

The TOS shows good internal consistency values for both subscales, "Healthy Orthorexia" (Cronbach's alpha=.80 to α=.87) and "Orthorexia Nervosa" (α=.81 to α=.90). The re-test reliability after 18 months was r>.70 for both subscales, according to one study. Even though a 2-factorial structure is theoretically meaningful, a 4-factorial solution and cross-loadings between the two subscales need to be investigated in future studies based on previous findings. Correlations with tools for disordered eating were positive for both subscales (moderate correlations for "Healthy Orthorexia" and moderate to high for "Orthorexia Nervosa"), unless the other subscale was partialled out, which reversed and attenuated the correlational relationship for "Healthy Orthorexia" to "Bulimia" (EAT-26) and other subscales measuring disordered eating, negative affect and perfectionism. Partialling out "Healthy Orthorexia", did not change the relationship between "Orthorexia Nervosa" and OCD-symptoms, disordered eating behaviors and perfectionism, but increased the positive correlation with negative affect.

4 Discussion

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

318 This systematic review collated all available information on the psychometric
319 properties of ten quantitative ON measures, in order to illustrate weaknesses and
320 strengths of existing ON tools. The review had two key findings. First, only 68 out of
321 the 141 (48.23%) identified studies reported at least one of the specified
322 psychometric quality assessments. This finding is surprising, as there is no 'gold
323 standard' or commonly accepted definition of ON (e.g. Cena et al., 2018)
324 complicating any validity judgements. The lack of exhaustive reliability analyses
325 suggests that prevalence rates and ON-risk assessments could be highly affected by
326 measurement errors. Reporting different reliability scores is important, as they
327 represent different cumulative sources of measurement error (e.g. Henson, 2001).
328 More transparency regarding measurement properties should be displayed, as they
329 are essential for researchers choosing their measures.

330 The second finding of this review concerns the measures' reported psychometric
331 properties. Based on the present analysis, utilizing the BOT or the ORTO for the
332 assessment of ON is discouraged. This finding is in line with previous studies
333 advising against the use of the ORTO-15 (Dunn & Bratman, 2016; Missbach, Dunn &
334 König, 2017; Cena et al., 2018) and BOT (Eriksson et al., 2008) for assessing ON.
335 The original questions of the BOT were designed as a personal risk assessment,
336 which were updated by Bratman to become the 'The Authorized Bratman Orthorexia
337 Self-Test' (Bratman, 2017a). However, the original questions were never intended to
338 be used as an assessment tool (Bratman, 2017), as the BOT has no test-theoretical
339 foundation, which is reflected in its questionable dimensionality. It is not clear how
340 many of the BOT items refer to ON-specific behaviors and thoughts or general eating
341 pathology. The ORTO-15 had to be adapted by multiple studies to obtain acceptable
342 psychometric properties, in most cases by means of excluding a considerable

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

number of items. Furthermore, the inconsistent findings regarding the ORTO's construct validity suggest either problems with the conceptualization of the ORTO and/or common misinterpretations of score results.

Promising findings were identified for the DOS, the EHQ-R and the TOS, even though further validation in various (cultural) contexts are needed. For example, the DOS shows good reliability, however, most of its evaluations were carried out by the scale's author within German study samples and need to be replicated in other settings. Furthermore, the scale's dimensionality remains inconclusive, with a single factor being meaningful but mostly poorly fitted.

The B-ORA and the ONS were both designed and published in the framework of a thesis and will require further evaluation. Finally, the Puerto Rican tool and the BISQ ON-subscale were designed for specific cultural contexts. At this point, it is not recommended to apply these scales without further validation.

Some of the measurement difficulties identified may reflect a lack of conceptual clarity regarding ON more generally. The findings regarding the DOS's high correlations with measures for weight and shape concerns deviate from the common perception that ON is not related to an intentional weight loss (e.g. Dunn & Bratman, 2016). Moreover, it remains to be clarified what constitutes a pathological approach to healthy eating. Feelings of superiority regarding one's healthy diet, for example, were originally seen as a core element of the clinical picture (Bratman & Knight, 2000) describing ON. Within the TOS, items related to feelings of superiority are conversely seen as part of "Healthy Orthorexia".

This systematic review identified all measures published at the time of the analysis. New measures are continuously being developed, such as the Barcelona Orthorexia

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

Scale (BOS, Bauer et al., 2018), which is based on a Delphi study methodology, and the Orthorexia Nervosa Inventory (ONI, Oberle, De Nadai & Madrid, 2020). The ONI combines adapted items from the DOS and EHQ with novel items to represent coinciding diagnostic criteria of ON. A first validation study indicated high internal consistency values (Cronbach's alpha=.94 for ONI_{total}, α>.88 for all subscales) and high positive correlations with the EAT-26 (r=.79 for ONI_{total}) and the Yale-Brown Obsessive-Compulsive Scale (r=.53). The new scale construction approaches are promising as they are combining various expert opinions. However, this review has shown that there are conceptual differences among ON scales that represent disparate expert opinions and need to be addressed. Evaluating and comparing ten ON scales revealed the individual strengths and weaknesses of the measures and will provide guidance for future research within the field.

5 Strengths and Limitations

Even though past reviews have scrutinized ON research, no study so far exhaustively evaluated the psychometric properties of all ON assessment tools. The present systematic review included all empirical studies assessing ON, which were published in English, German, French, Dutch or Spanish up to the end of August 2019. Even though articles written in multiple languages were included in this review, twelve studies were not assessed due to being written in Portuguese, Italian, Polish or Swedish. Excluding these articles might have introduced a language bias considering that the presented results varied across different language versions.

The RG analysis in this systematic review could have been influenced by the so-called "file drawer problem" (Rosenthal, 1979; Howell & Shields, 2008). Only about half of the studies that assessed ON with a quantitative measure reported score

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

reliabilities. This problem was partly addressed by contacting authors via email if any information was missing or ambiguous, which resulted in additional information being provided by 16 authors.

The studies included in this review were heterogeneous in many respects, including their conceptualization of ON, their assessment method and their study samples. This heterogeneity made a comparison difficult and results should be interpreted bearing in mind that study designs and approaches greatly varied between studies. The goal of this review was not to evaluate the quality of included studies' methodological approaches, but rather contrasting the information provided regarding reported psychometric properties.

6 Conclusion

The analysis of reliability and validity indicators of ON measures demonstrated that existing tools exhibit either questionable psychometric properties (BOT, ORTO-15), challenge preliminary diagnostic criteria (DOS, TOS) or require further evaluation (e.g. EHQ-R, ONS, B-ORA). A surprising and concerning finding of this review refers to the lack of reporting in relation to psychometric properties, considering that no gold standard exists for the measurement of ON. Further research is needed to clarify current inconsistencies in the conceptualization of ON, which are reflected in its measurement tools. Additionally, future studies need to be more transparent about the process of test construction and evaluation, if we want to improve the research surrounding ON. Potential implications of this analysis are therefore the need for more rigorous evaluation processes for new and existing scales across

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

415 settings and cultural contexts, recognizing the provisional nature of any research
416 findings associated with scales intending to measure ON and the necessity of
417 researchers and practitioners to address current contradictions in the
418 conceptualization of ON.

419 **Acknowledgement**

420 We would like to thank Dr Friederike Barthels and Tobias Thomas for providing us
421 with unpublished data. Furthermore, we would like to thank Dr David Gleaves and Dr
422 Radoslaw Rogoza for providing unpublished material to include in our analysis.

423 **Author Contributions**

424 All authors have approved the final article for publication. MCO, HS and EN
425 designed the study, MCO, ASAVM and MR conducted the searches, MCO drafted
426 the manuscript, HS and EN provided revisions to the manuscript.

427 **Disclosure**

428 The authors of this review have no conflict of interest to declare. Ethical approval
429 was not required, as this was a review study.

430 **Funding Sources**

431 This research did not receive any specific grant from funding agencies in the public,
432 commercial, or not-for-profit sectors.

433 **References**

434 American Psychiatric Association. (2013). *Diagnostic and statistical manual of*
435 *mental disorders* (5th ed.). Washington, DC.

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

436 Andreas, S., Schedler, K., Schulz, H., & Nutzinger, D. (2018). Evaluation of a
437 German version of a brief diagnosis questionnaire of symptoms of orthorexia
438 nervosa in patients with mental disorders (Ortho-10). *Eating and Weight*
439 *Disorders - Studies on Anorexia, Bulimia and Obesity, 23(1*), 75-85.

440 Arusoğlu, G., Kabakçi, E., Köksal, G., & Merdol, T. K. (2008). Orthorexia Nervosa
441 and Adaptation of ORTO-11 into Turkish. *Turkish journal of psychiatry*, *19*(3).

442 Barrada, J. R., & Roncero, M. (2018). Bidimensional structure of the orthorexia:
443 Development and initial validation of a new instrument. *Anales De*
444 *Psicología/Annals of Psychology*, *34*(2), 283-291.

445 Barthels, F., Meyer, F., & Pietrowsky, R. (2015a). Die Düsseldorf Orthorexie
446 Skala–Konstruktion und Evaluation eines Fragebogens zur Erfassung ortho-
447 rektischen Ernährungsverhaltens. Zeitschrift für Klinische Psychologie und
448 Psychotherapie. *Zeitschrift für Klinische Psychologie Psychotherapie, 44(2).*
449 97–105.

450 Barthels, F., Meyer, F., & Pietrowsky, R. (2015b). Orthorexic eating behavior.
451 *Ernährungsumschau*, *62*(10).

452 Bauer, S., Fusté, M., Andrés, A., & Saldaña, A. (2018). The Barcelona
453 Orthorexia Scale (BOS): Development process using the Delphi method.
454 *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity,*
455 *24(2),* 247-255.

456 Bland, J., & Altman, D. (1997). Statistics notes: Cronbach's alpha. *BMJ, 314*(7080),
457 572-572.

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

458 Booth, A. (2012). The pros and pros of registration on PROSPERO. *BJOG: An*
459 *International Journal of Obstetrics and Gynaecology, 119*(8), 904-905.

460 Bratman, S. (1997). Health Food Junkie*. Yoga Journal,* 42-50.

461 Bratman, S. (2017a). The Authorized Bratman Orthorexia Self-Test. Retrieved from
462 http://www.orthorexia.com/the-authorized-bratman-orthorexia-self-test/

463 Bratman, S. (2017b). Orthorexia vs. theories of healthy eating. *Eating and Weight*
464 *Disorders*, *22*(3), 381–385.

465 Bratman, S. & Knight, D. (2000). *Health Food Junkies: Overcoming the Obsession*
466 *with Healthful Eating.* New York: Broadway Books.

467 Brytek-Matera, A., Gramaglia, C., & Zeppegno, P. (2014a). Healthy eating obsession
468 in women with anorexia nervosa: a case control study. *New developments in*
469 *anorexia nervosa research. New York: Nova*, 39-50.

470 Brytek-Matera, A., Krupa, M., Poggiogalle, E., & Donini, L. M. (2014b). Adaptation of
471 the ORTHO-15 test to Polish women and men. *Eating and Weight Disorders-*
472 *Studies on Anorexia, Bulimia and Obesity*, *19*(1), 69-76.

473 Burda, L. (2018). *Development and Validation of an Inventory Measuring Dietary*
474 *Attitudes of Healthy Eating and Orthorexia Nervosa,* ProQuest Dissertations
475 and Theses.

476 Carrero, T., Padilla Cotto, L., & Rodríguez-Gómez, J. (2016). Construcción y
477 Validación de una Escala para Tamizar Ortorexia: Primer estudio piloto en
478 Puerto Rico. *Boletin De La Asociacion Medica De Puerto Rico, 108*(2), 89-95.

479 Cena, H., Barthels, F., Cuzzolaro, M., Bratman, S., Brytek-Matera, A., Dunn, T., . . .
480 Donini, L. (2018). Definition and diagnostic criteria for orthorexia nervosa: A

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

narrative review of the literature. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity, 24*(2), 209-246.

Chard, C., Hilzendegen, A., Barthels, C., & Stroebele-Benschop, F. (2018). Psychometric evaluation of the English version of the Düsseldorf Orthorexie Scale (DOS) and the prevalence of orthorexia nervosa among a U.S. student sample. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity, 24*(2), 275-281.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement, 20*(1), 37-46.

Costa, C., Hardan-Khalil, K., & Gibbs, K. (2017). Orthorexia Nervosa: A Review of the Literature. *Issues in Mental Health Nursing, 38*(12), 980-988.

DeVellis, R. (2003). *Scale development: Theory and applications* (Second ed., Applied social research methods series; v. 26). Thousand Oaks, California; London: Sage Publications.

Donini, L., Marsili, M., Graziani, D., Imbriale, P., & Cannella, C. (2004). Orthorexia nervosa: A preliminary study with a proposal for diagnosis and an attempt to measure the dimension of the phenomenon. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity, 9*(2), 151-157.

Donini, L., Marsili, M., Graziani, D., Imbriale, P., & Cannella, C. (2005). Orthorexia nervosa: Validation of a diagnosis questionnaire. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity, 10*(2), E28-E32.

Dunn, T., & Bratman, S. (2016). On orthorexia nervosa: A review of the literature and proposed diagnostic criteria. *Eating Behaviors, 21*, 11-17.

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

Eriksson, L., Baigi, A., Marklund, B., & Lindgren, E. C. (2008). Social physique anxiety and sociocultural attitudes toward appearance impact on orthorexia test in fitness participants. *Scandinavian journal of medicine & science in sports*, *18*(3), 389-394.

Gleaves, D. H., Graham, E. C., & Ambwani, S. (2013). Measuring "orthorexia": Development of the Eating Habits Questionnaire. *The International Journal of Educational and Psychological Assessment 12(2),* 1-18.

Glen, K., & Gleaves, D. (2018). *Development and Validation of the Eating Habits Questionnaire-Revised (EHQ-R)*. (Unpublished master thesis). University of South Australia, Adelaide, South Australia.

Graham, E. C. (2005). *Development of the eating habits questionnaire* (Doctoral dissertation, Texas A&M University).

Grammatikopoulou, M., Gkiouras, G., Markaki, K., Theodoridis, A., Tsakiri, X., Mavridis, V., . . . Chourdakis, T. (2018). Food addiction, orthorexia, and food-related stress among dietetics students. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity, 23*(4), 459-467.

Haddad, Chadia, Hallit, Rabih, Akel, Marwan, Honein, Karl, Akiki, Maria, Kheir, Nelly, . . . Hallit, Souheil. (2019). Validation of the Arabic version of the ORTO-15 questionnaire in a sample of the Lebanese population. *Eating and Weight Disorders: EWD,* Eating and weight disorders: EWD, 22 May 2019.

Hakstian, A. R., & Whalen, T. E. (1976). A k-sample significance test for independent alpha coefficients. *Psychometrika*, *41*(2), 219-231.

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

He, J., Ma, H., Barthels, F., & Fan, X. (n.d.). Psychometric properties of the Chinese version of the Düsseldorf Orthorexia Scale: Prevalence and demographic correlates of orthorexia nervosa among Chinese university students. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity, 24*(3), 453-463.

Heiss, S., Coffino, J., & Hormes, J. (2019). What does the ORTO-15 measure? Assessing the construct validity of a common orthorexia nervosa questionnaire in a meat avoiding sample. *Appetite, 135*, 93-99.

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and evaluation in counseling and development*, *34*(3), 177-189.

Higgins J.P.T., Thomas J., Chandler J., Cumpston M., Li T., Page M.J., Welch V.A. (Eds). (2019). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.0 (updated July 2019). Cochrane. Available from www.training.cochrane.org/handbook.

Howell, R., & Shields, A. (2008). The File Drawer Problem in Reliability Generalization: A Strategy to Compute a Fail-Safe N With Reliability Coefficients. *Educational and Psychological Measurement, 68*(1), 120-128.

Jenaro Río, C., Flores Robaina, N., Bermejo, B. G., & Cruz, M. I. (2011). Cuestionario de Imagen Corporal para la detección temprana de trastornos de la conducta alimentaria. *Acción Psicológica, 8*(1), 7-20.

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

Kinzl, J. F., Hauer, K., Traweger, C., & Kiefer, I. (2005). Orthorexia Nervosa: Eine Häufige Essstörung bei Diätassistentinnen. *Ernährungs-Umschau*, *11*, 436-439.

Koven, N., & Abry, A. (2015). The clinical basis of orthorexia nervosa: Emerging perspectives. *Neuropsychiatric Disease and Treatment, 11*, 385-394.

Kramer, M. (2016). *The Development of a Scale to Measure Orthorexia Nervosa,* ProQuest Dissertations and Theses.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Missbach, B., Dunn, T., & König, J. (2017). We need new tools to assess Orthorexia Nervosa. A commentary on “Prevalence of Orthorexia Nervosa among College Students Based on Bratman’s Test and Associated Tendencies”. *Appetite, 108*, 521-524.

Missbach, B., Hinterbuchinger, B., Dreiseitl, V., Zellhofer, S., Kurz, C., König, J., & Manalo, E. (2015). When Eating Right, Is Measured Wrong! A Validation and Critical Examination of the ORTO-15 Questionnaire in German (Orthorexia Nervosa Validation in German). *10*(8), E0135772.

Mokkink, L., De Vet, H., Prinsen, C., Patrick, D., Alonso Caballero, J., Bouter, L., & Terwee, C. (2018). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *27*(5), 1171-1179.

Moroze, R., Dunn, T., Craig Holland, J., Yager, J., & Weintraub, P. (2015). Microthinking About Micronutrients: A Case of Transition From Obsessions

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

About Healthy Eating to Near-Fatal “Orthorexia Nervosa” and Proposed Diagnostic Criteria. *Psychosomatics, 56*(4), 397-403.

Novara, C., Pardini, S., Pastore, M., & Mulatti, C. (2017). Ortoressia Nervosa: un’indagine del costrutto e delle caratteristiche psicometriche della versione italiana dell’Eating Habits Questionnaire-21 (EHQ-21). *Psicoterapia Cognitiva e Comportamentale*, *23*(3).

Nunnally, J. (1967). *Psychometric theory* (McGraw-Hill series in psychology). New York: McGraw-Hill.

Oberle, C. D., De Nadai, A. S., & Madrid, A. L. (2020). Orthorexia Nervosa Inventory (ONI): development and validation of a new measure of orthorexic symptomatology. *Eating and Weight Disorders: EWD*.

Parra-Fernández, M. L., Onieva-Zafra, M. D., Fernández-Muñoz, J. J., & Fernandez-Martinez, E. (2019). Adaptation and validation of the Spanish version of the DOS questionnaire for the detection of orthorexic nervosa behavior. *PloS one*, *14*(5).

Parra-Fernandez, M., Rodriguez-Cano, T., Onieva-Zafra, M., Perez-Haro, M., Casero-Alonso, V., Munoz Camargo, J., & Notario-Pacheco, B. (2018). Adaptation and validation of the Spanish version of the ORTO-15 questionnaire for the diagnosis of orthorexia nervosa. *PLoS ONE, 13*(1), E0190722.

Prinsen, C., Mokkink, A., Bouter, L., Alonso, B., Patrick, M., Vet, J., & Terwee, D. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*(5), 1147-1157.

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

592 Rodriguez, M., & Maeda, Y. (2006). Meta-Analysis of Coefficient Alpha.
593 *Psychological Methods, 11*(3), 306-322.

594 Rogoza, R. (2019). [untitled manuscript]. (Unpublished manuscript).

595 Rosenthal, R. (1979). The file drawer problem and tolerance for null results.
596 *Psychological bulletin*, *86*(3), 638.

597 Setnick, J. (2013). The eating disorders clinical pocket guide, second edition: quick
598 reference for healthcare providers. *Understanding Nutrition*, Dallas.

599 Sideri, S., Papageorgiou, S., & Eliades, T. (2018). Registration in the international
600 prospective register of systematic reviews (PROSPERO) of systematic review
601 protocols was associated with increased review quality. *Journal of Clinical*
602 *Epidemiology, 100*, 103-110.

603 Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International*
604 *Journal of Medical Education, 2*, 53-55.

605 Terwee, Bot, De Boer, Van Der Windt, Knol, Dekker, . . . De Vet. (2007). Quality
606 criteria were proposed for measurement properties of health status
607 questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34-42.

608 Terwee, C., Prinsen, B., Chiarotto, A., Westerman, M., Patrick, J., Alonso, D., . . .
609 Mokkink, W. (2018). COSMIN methodology for evaluating the content validity
610 of patient-reported outcome measures: A Delphi study. *Quality of Life*
611 *Research, 27*(5), 1159-1170.

612 Valente, M., Syurina, E., & Donini, V. (2019). Shedding light upon various tools to
613 assess orthorexia nervosa: A critical literature review with a systematic

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

search. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity, 24*(4), 671-682.

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), 59.

World Health Organization. (2018). *International classification of diseases for mortality and morbidity statistics* (11th Revision). Retrieved from https://icd.who.int/browse11/l-m/e

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

Table 1 Identified ON Measures

Tool	Author(s) and Year	Country	Conceptual Framework	Measurement Aim	Format and Practicalities	Target Population	Item Selection Strategy	Language Adaptations
Body Image Screening Questionnaire (BISQ)	Jenaro Río et al. (2011)	Spain	Conceptualizing eating disorders from a comprehensive perspective, which considers ON and Vigorexia as part of the eating disorder spectrum	Designing an early detection tool for different eating disordered behaviors	24-items (total scale); 5 items to measure ON, scored on a 6-point Likert scale	Evaluated within a population of participants considered as at-risk for disordered eating behaviors and a general population sample	Clinical criteria to identify ON evaluated by four experts in the field of eating disorders	---
Burda Orthorexia Risk Assessment (B-ORA)	Burda (2018)	United States	Exploring cognitive, emotional, and behavioral themes leading to problematic ON-related eating habits	Not a diagnostic tool; providing information for clinicians to detect unhealthy eating behaviors and phobic-obsessive personality traits associated with ON	27 items (21 ON-related tendencies, 6 control questions, which are not scored; 4-point Likert scale ('I strongly disagree' to 'I strongly agree')	College students within the United States	Items based on proposed diagnostic criteria by Dunn & Bratman (2016) and Moroze et al. (2015), as well as ON-specific patterns of thoughts, emotions and behaviors as presented in the literature	---
Bratman Orthorexia Test (BOT)	Based on the informal personal	United States	ON as a fixation on/obsession	Identifying potential	10 dichotomous questions, which can be	People being overly	The content of the self-assessment	German (Kinzl et al. (2005), Swedish (Eriksson et al.,

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

	risk-assessment quiz originally published in "Health Food Junkies" (Bratman & Knight, 2000)		with eating healthy food	problems with food habits	answered with either 'yes' or 'no'; scoring differs between authors	concerned with healthy nutrition	questions are based on the personal as well as professional experiences of Steven Bratman, who first coined the term ON in 1997	2008) Greek (Grammatikopoulou et al., 2018), with differing score interpretations
Düsseldorf Orthorexia Scale (DOS)	Barthels, Meyer & Pietrowsky (2015a)	Germany	Measuring a potentially pathological fixation on a health-conscious diet, considering cultural particularities	Assessing orthorexic eating behaviors with discriminatory power, based on test-theoretical considerations	10-items; 4-point Likert scale ('does not apply to me, 'rather does not apply to me' to 'rather applies to me' and 'applies to me'); preliminary cut-off at 30 points	General German Population	Inductive item generation involving relevant eating behaviors and attitudes towards nutritional knowledge, taking into account the case studies in Bratman and Knight (2000).	English (Chard et al., 2018), Chinese (He et al., 2019), Spanish (Parra-Fernandez et al., 2019)
Eating Habits Questionnaire (EHQ)	Graham (2005); Gleaves et al. (2013)	United States	ON as "[…] an otherwise healthy behavior […] taken to extremes […]" (Gleaves et al., 2013, p. 1).	Measuring a pathological fixation on eating healthy food	21-item scale; 4-point Likert scale ('false, not at all true' to 'very true'); three subscales (problems associated with healthy eating, knowledge of healthy eating and feeling positively about healthy eating);	Validated with a population of college students studying psychology or nutrition	The original item pool was generated according to the information provided in the book "Health Food Junkies" and evaluated by four graduate students trained in ON symptomatology to select items	Italian (Novara et al., 2017; not included in this review as only available in Italian)

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

					higher values indicate more ON symptoms		accurately representing ON	
Eating Habits Questionnaire-Revised (EHQ-R)	Glen & Gleaves (2018)	Australia	Considering impairments and negative emotionality associated with ON; including the proposed diagnostic criteria by Dunn & Bratman (2016)	To extend and update the EHQ in order to identify the underlying factors describing ON	30 items; five factors: rigidity (7 items), healthy body image (6 items), violation of dietary rules (7 items), negative emotionality (6 items), time impairment (4 items)	General Australian Population	25/30 based on Dunn & Bratman (2016) criteria and ON research; 5 items from the EHQ	---
Orthorexia Nervosa Scale (ONS)	Kramer (2016)	United States	ON as associated with superiority, downward social comparison, rigidity, purity, social avoidance, identity, eating disorder as meaning, loss of control, preoccupation, eating to cope, nutritional deficiencies and relationship problems	Assessing current behavioral tendencies associated with ON	47 items with 10 dimensions; 5-point Likert-scale ("Strongly Disagree", "Disagree", "Neither Agree or Disagree", "Agree", "Strongly Agree")	Developed and evaluated with a college student sample	Items based on previous questionnaires (ORTO-15, EHQ, BOT) and a literature review on ON combined with the consultation of colleagues familiar with the content area	---
ORTO-15	Donini et al. (2004; 2005)	Italy	ON as a "[..] more or less serious personality or behavioral	Measuring cognitive-rational, clinical and emotional aspects of ON	15 items scale with a closed multiple-choice format ('always', 'often',	Italian general population	6/15 items based on questions from Bratman's Orthorexia self-	Turkish (Arusoglu et al., 2008), Polish (Brytek-Matera et al., 2014b), Hungarian (Varga

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

			disturbance […]" (Donini et al., 2004, p. 151).		'sometimes', 'never'). Higher values indicate healthier behaviors. For items 1 and 13, the highest values are assigned to the answers 'often' (4) and 'sometimes' (3); cut-off <40 points to indicate a potential diagnosis		test, with some of the wordings changed; preliminary versions of the questionnaire were piloted to create the final measurement tool	et al., 2014), German (Missbach et al., 2015), Spanish (Parra-Fernandez et al., 2018), English (Heiss, Coffino & Hormes, 2019) and Arabic (Haddad et al., 2019)
Scale to Measure Orthorexia in Puerto Rican Men and Women	Carrero, Cotto & Rodriguez-Gomez (2016)	Puerto Rico	Designing a measure for ON, which is applicable in the Latin American context; ON defined according to Bratman (1997)	Assessing ON within a Latin American population including behaviors, thoughts, feelings, perceptions and symptomatology	The five subscales "Food Intake", "Obsession", "Lifestyle", "Social Isolation" and "Pollution and Compulsion" constitute the measure. The 27-item tool uses a 4-point Likert scale ('never' to 'always'), with higher values representing a higher ON symptomatology	Puerto Rican men and women	Items are based on ON research and an expert rating, including raters with experience in psychology, in the field of eating disorders and in evidence building	---

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

Teruel Orthorexia Scale (TOS)	Barrada & Roncero, (2018)	Spain	Orthorexia seen as either a tendency and interest in eating healthy food (HeOr) or a pathological preoccupation with a rigid healthy diet (OrNe)	Measuring both problematic and non-problematic healthy eating behaviors and differentiating between the two	17 items measuring two dimensions: "Healthy Orthorexia" (HeOr, 9 items) and "Orthorexia Nervosa" (OrNe, 8 items); 4-point rating scale ('completely disagree' to 'completely agree'); higher values indicate a higher expression of either HeOr or OrNe.	Validated with primarily university student sample	An initial item pool for the scale was generated to represent the previous literature on ON, which was then reduced based on the principle of statistical simplicity	---

Table 2 Evaluation of Psychometric Properties

Tool	**Scales**	**Internal Consistency**		**Re-test Reliability**	**Construct Validity**	**Structural Validity***
BISQ	ON-subscale	5 items	Cronbach's alpha=.59-.77 **(k=2)**	---	---	5 factors for total scale (ON one factor with 5 items) **(k=1)**

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

B-ORA	Total	21 items	Cronbach's alpha=.97 **(k=1)**	---	EAT-26: r=.45 **(k=1)**	Single Factor (with adjusted error terms) **(k=1)**
BOT	Total	10 items	Cronbach's alpha=.67-.79; KR-20= .60 **(k=3)**	---	FEV: Those at risk for ON (≥4 affirmative answers) also showed a statistically higher "Restraint" and "Disinhibition", as measured by the FEV (no effect sizes reported) **(k=1)**	---
		9 Items	Cronbach's alpha =.73-.77 **(k=1)**	---	**Total scale:** EDI-2: r=.53; "Drive for Thinness" (r=.59), "Interoceptive Awareness" (r=.49) and "Asceticism" (r=.48) **ON-specific subscale:** EDI-2: r=.28 **(k=1)**	2-factor solution **(k=1)**
DOS	Total	10 items	Cronbach's alpha=.69-.93 **(k=13)**	**3 times every 3 months:** Cronbach's alpha=.79 t1, α=.84 t2 and α=.83 t3 (t1 and t2: r=.67, t1 and t3: r=.73; t2 and t3: r=.79) **4 weeks:** r=.77 **(k=2)**	WI: r=.24 EDI: r=.13 ("Interpersonal Distrust") to r=.48 ("Drive for Thinness") EDI-2: r=.53, r=.54, r=.50 ("Drive for Thinness"), r=.32, r=.41, r=.20 ("Bulimia"), r=.27, r=.30 ("Body Dissatisfaction"), r=.48, r=.33 ("Asceticism"), r=.18 ("Perfectionism"), r=.37, r=.24 ("Interoceptive Awareness"), r=.20 ("Impulse Regulation"), r=.10 ("Social Insecurity"), r=.16 ("Effectiveness") FEV: r=.49 ("Cognitive Control") IEQ: r=.59 TFEQ-R18: r=.06 ("Cognitive Restrain"), r=-.10 ("Uncontrolled Eating") **(k=5)**	Strong principal component (poor model fit); 5 factors indicated (single factor better representation of construct, but poorly fitted); 3 factors; Single factor **(k=4)**
	Obsession in Healthy Food	5 items	Cronbach's alpha=.77 **(k=1)**	**4 weeks**: r=.71 **(k=1)**	IEQ: r=.50 TFEQ-R18: r=-.14 ("Uncontrolled Eating"), r=-.06 ("Emotional Eating") **(k=1)**	
	Adherence to strict nutrition	3 items	Cronbach's alpha=.75 **(k=1)**	**4 weeks:** r=.46 **(k=1)**	IEQ: r=.41 TFEQ-R18: r=.07 ("Emotional Eating") **(k=1)**	

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

	Emotional Symptoms	2 items	Cronbach's alpha=.71 **(k=1)**	**4 weeks:** r=.50 **(k=1)**	IEQ: r=.50 TFEQ-R18: r=.08 ("Cognitive Restrain"), r=-.08 ("Uncontrolled Eating") **(k=1)**	
EHQ	Total	21 items	Cronbach's alpha=.86-94 **(k=6)**	---	EDI-3: r=.28 for "Interpersonal Alienation" PROMIS-Anxiety: r=.19 OCI-R: r=.37 MOCI: r=.32 EAT-26: r=.79; r=.63; r=.56 EAT-26-SRT/BP: r=.37 BULIT-R: r=.62; r=.29 CES-D: r=.22 MEADS: r=.15 up to r=.69 NPI: r=.17 MPS: r=.23 **(k=5)**	3-factors with underlying general factor; 3 factors (with 3 items loading on a different factor); 3 factors with covarying residuals; 5 factors **(k=4)**
	Problems	12 items	Cronbach's alpha =.82-.92 **(k=4)**	**2-4 weeks:** r=.81 **(k=1)**	CES-D: r=.22; r=.30 PAS: r=.21 IPIP-41-N: r=.31 IPIP-41-C: r=.07 BIDR SDE: r=-.18 EAT-26: r=.79; r=.51; r=.67 (subscales) BULIT-R: r=.43 ("Body Image/Weight Loss"), r=.56 ("Vomiting/Laxatives) MEADS: r=.21 up to r=.61 → **Partialling out the other EHQ subscales:** EAT-26-SRT/BP: r=.44 NIAS: r=.28 up to r=.44 CIA-R: r=.30 OCI-R: r=.18 **(k=3)**	
		9 items	Cronbach's alpha=.71-.79 **(k=2)**	---	NPI: r=.11 MPS: r=.30; r=.13 up to r=.25 (subscales) **(k=1)**	
	Knowledge/ Behaviors	5 items	Cronbach's alpha =.81-.87 **(k=4)**	**2-4 weeks:** r=81 **(k=1)**	EAT-26: r=.54; r=.20 up to r=.27 (subscales) BULIT-R: r=.25 ("Vomiting/Laxatives) MAEDS: r=.56 ("Avoidance"), r=.23 ("Fear of Fatness"), r=.18 ("Purgative")	

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

					(k=2)	
		8 items	Cronbach's alpha =.86-.87 **(k=2)**	---	SES: r=.15 NPI: r=.18 MPS: r=.13 (r=.25 "Personal Standards", r=.11 "Organization") **(k=1)**	
	Feelings	4 items	Cronbach's alpha =.70-.86 **(k=5)**	**2-4 weeks:** r=.72 **(k=1)**	EAT-26: r=.41; r=.23 up to r=.45 (subscales) BULIT-R: r=.25 ("Body Image/Weight Loss"), r=.33 ("Vomiting/Laxatives) MAEDS: (all subscales except for "Depression") r=.19 up to r=.57 NPI: r=.11 MPS: r=.13 (r=.20 "Personal Standards", r=.20 "Organization") **(k=3)**	
EHQ-R	Total	30 items	Cronbach's alpha=.81-.96 **(k=3)**	---	EAT-26: r=.78 MEADS: r=.54 ("Binge Eating"), r=.53 ("Purgative Behavior"), r=.69 ("Fear of Fatness"), r=.63 ("Avoidance of Forbidden Foods"), r=.45 ("Depression") (**k=1)**	5 factors **(k=1)**
	Rigidity	7 items	Cronbach's alpha=.81-.89 **(k=3)**	---	EAT-26 (original scoring): r=.26-r=.51 EAT-26 (alternative scoring): r=.35-r=.52 MEADS: r=.19-r=.65 **(k=1)**	
	Healthy Body Image	6 items	Cronbach's alpha=.82-.86 **(k=3)**	---	EAT-26 (original scoring): r=.18-r=.56 EAT-26 (alternative scoring): r=.25-r=.64 MEADS: r=.29-r=.65 **(k=1)**	
	Violation of Dietary Rules	7 items	Cronbach's alpha=.83-.89 **(k=3)**	---	EAT-26 (original scoring): r=.41-r=.75 EAT-26 (alternative scoring): r=.51-r=.75 MEADS: r=.47-r=.65 **(k=1)**	
	Negative Emotionality	6 items	Cronbach's alpha=.83-.89 **(k=3)**	---	EAT-26 (original scoring): r=.29-r=.67 EAT-26 (alternative scoring): r=.38-r=.69 MEADS: r=.46-r=.64 **(k=1)**	
	Time Impairment	4 items	Cronbach's alpha=.75-.86	---	EAT-26 (original scoring): r=.29-r=.70 EAT-26 (alternative scoring): r=.37-r=.72	

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

			(k=3)		MEADS: r=.44-r=.54 **(k=1)**	
ONS	Social/Interpersonal Concerns	10 items	Cronbach's alpha=.92 **(k=1)**	---	YFAS: r=.15 **(k=1)**	10 factors (loading on second order factor) **(k=1)**
	Discipline/Control	9 items	Cronbach's alpha=.90 **(k=1)**	---		
	Superiority/ Knowledge	5 items	Cronbach's alpha=.84 **(k=1)**	---		
	Pureness/Natural Quality	6 items	Cronbach's alpha=.81 **(k=1)**	---		
	Detox/Restricting	5 items	Cronbach's alpha=.75 **(k=1)**	---		
	Nutritional Deficiencies	2 items	Cronbach's alpha=.89 **(k=1)**	---		
	Online Forums/Blogs	3 items	Cronbach's alpha=.70 **(k=1)**	---		
	Defensiveness	2 items	Cronbach's alpha=.83 **(k=1)**	---		
	Fulfilment/Peace	2 items	Cronbach's alpha=.83 **(k=1)**	---		
	Fasting	3 items	Cronbach's alpha=.69 **(k=1)**	---		
ORTO-15	Total	15 items	Cronbach's alpha= .14-.84 **(k=28)**	**2 weeks:** Kappa=.66-.89 **(k=1)**	EAT-26: r=-.12; r=-.22 r=-.33; r=-.51 EAT-26 (ON-specific): r=-.18 EDI-2: r=-.14 EDI-3: r=-.32 "Interpersonal Alienation", r=-.31 "Perfectionism", r=-.22 "Asceticism" DRES: r=-.20	3 factors **(k=1)**

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

			OCI-R: r=-.21 EDE-Q: meeting the cut-off not related to meeting ORTO-15 cut-off (35 or 40); r=-.13 to r=-.34 with "Restraint" (for different diets) BULIT-R: r=-.36 MEADS: r=-.15 up to r=-.44 (ORTO-15) and CES-D: r=-.19 **(k=6)**	
13 items	---	---	EDE-Q: Stepwise Regression analysis: disordered eating patterns were the only significant predictor for factor 1 (β=-.26) and factor 2 (β=-.28) of the ORTO-15 (interpreted as a higher level of disordered eating patterns related to fewer ON symptoms) **(k=1)**	2 factors **(k=1)**
12 items	Cronbach's alpha=.39 **(k=1)**	---	EAT-26: r=-.26 EAT-26 (ON specific): r=-.23 **(k=1)**	3 factors **(k=2)**
11 items	Cronbach's alpha=.62-.86 **(k=10)**	30 days: Cohen's Kappa for individual items: .91-1.00 **(k=1)**	EAT-26: r=-.26, r=-.26, r=-.28 EAT-26 (ON specific): r=-.17, r=-.16, r=-.24 EAT-40: A one-way ANOVA showed that eating attitude had a significant main effect on orthorexic tendency (F (2,993) = 48.04 p<.001); Turkey's test indicated that those with pathological eating attitudes also had a higher orthorexic tendency MOCI: A one-way ANOVA showed that obsessive-compulsive symptoms had a significant main effect on orthorexic tendency (F (2,993) = 27.56; p< .001); Turkey's test indicated that individuals with higher obsessive-compulsive symptoms displayed equally greater orthorexic tendencies **(k=2)**	Single factor; 3 factors **(k=4)**
9 items	Cronbach's alpha=.67-.74 **(k=3)**	---	EAT-26: r=-.37, r=-.34; r=-.28 with factor 1 of the ORTO-9 ("Dieting" r=-.36); factor 2: r=-.65 ("Dieting" r=-.59, "Bulimia and Food Preoccupation" r=-.67, "Oral Control" r=-.23)	2 factors Single factor **(k=2)**

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

					EAT-26 (ON specific): r=-.31, r=-.31 **(k=2)**	
		7 items	Cronbach's alpha=.83 (k=1)	---	---	Single factor **(k=1)**
		6 items	Cronbach's alpha=.79; McDonald's ω=.75 **(k=2)**	---	CES-D: r=-.45 EAT-26: r=-.74 BULIT-R: r=-.66 MAEDS: r=-.45 up to r=-.64 **(k=1)**	Single factor with latent factor for method bias **(k=1)**
The Puerto Rican ON Tool	Total	22 items	Cronbach's alpha=.87 **(k=1)**	---	---	5 factors **(k=1)**
	Food Intake	4 items	Cronbach's alpha=.84 **(k=1)**	---	---	
	Obsession	6 items	Cronbach's alpha=.78 **(k=1)**	---	---	
	Lifestyle	4 items	Cronbach's alpha=.66 **(k=1)**	---	---	
	Social Isolation	4 items	Cronbach's alpha=.77 **(k=1)**	---	---	
	Contamination and Compulsion	4 items	Cronbach's alpha=.74 **(k=1)**	---	---	
TOS	Healthy Orthorexia (HeOr)	9 items	Cronbach's alpha=.80-.87 **(k=4)**	**18 months:** r=.73 **(k=1)**	EAT-26: r=.30 ("Diet"), r=.22 ("Bulimia"), r=.22 ("Oral Control") MBSRQ: r=.11 ("Appearance Evaluation") → Partialling out OrNe: EAT-26: r=-.13 ("Bulimia"), r=.08 ("Oral Control") DEBQ: r=-.18/-.18 ("Restrained Eating"), r=-.22/.02 ("Emotional Eating"), r=-.25/.11 ("External Eating") PANAS: r=-.20, β=-.42 ("Negative Affect"); β=.27/.52 ("Positive Affect")	2 factors (theoretically more meaningful than 4); 2 factors (bi-dimensional structure with cross-loadings) **(k=2)**

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

				FMPS: r=-.13 (“Concern over Mistakes”) **(k=2)**
Orthorexia Nervosa (OrNe)	8 items	Cronbach’s alpha=.81-.90 **(k=4)**	**18 months:** r=.82 **(k=1)**	OCI-R: r=.32 EAT-26: r=.67 (“Diet”), r=.67 (“Bulimia”), r=.35 (“Oral Control”) DEBQ: r=.53/.60 (“Restrained Eating”), r=.24/.35 (“Emotional Eating”), r=-.08/.06 (“External Eating”) PANAS: r=.28, β=.26/.66 (“Negative Affect”); β=-.26/-.30 (“Positive Affect”) FMPS: r=.41 (“Concern over Mistakes”) → Partialling out HeOr: OCI-R: r=.33 EAT-26: r=.62 (“Diet”), r=.65 (“Bulimia”), r=.28 (“Oral Control”) PANAS: r=.34 (“Negative Affect”) FMPS: r=.42 (“Concern over Mistakes”), r=-.34 (“Appearance Evaluation”) (k=2)

* Factor structure as identified by study authors

BULIT-R=Bulimia Test-Revised, BIDR=Balanced Inventory of Desirable Responding, CES-D=Center for Epidemiologic Studies Depression Scale, CIA-E=Clinical Impairment Assessment-Eating, DEBQ=Dutch Eating Behavior Questionnaire, EAT=Eating Attitudes Test, EDE-Q=Eating Disorder Examination Questionnaire, EDI=Eating Disorder Inventory, FEV=Fragebogen zum Essverhalten (German version of the Three-Factor Eating Questionnaire), DRES=Dutch Restrained Eating Scale, FMPS=Frost Multidimensional Perfectionism Scale, HeOr=Healthy Orthorexia, IEQ=Inflexible Eating Questionnaire, IPIP=International Personality Item Pool (N for Neuroticism, C for Conscientiousness), KR-20=Kuder-Richardson-20, MAEDS=Multifactorial Assessment of Eating Disorders Symptoms, MBSRQ=Multidimensional Body-Self Relations Questionnaire, MOCI=Maudsley Obsessive-Compulsive Inventory, MPS=Multidimensional Perfectionism Scale, NIAS=Nine-Item-Avoidant/restrictive-food-intake-disorder-Screen, NPI=Narcissistic Personality Inventory, OCI-R=Obsessive-Compulsive Inventory-Revised, OrNe=Orthorexia Nervosa, PANAS=Positive-Affect-Negative-Affect-Scale, PAS=Personality Assessment Screener, PROMIS=Patient-Reported Outcomes Measurement Information System, SES=Rosenberg-Self-Esteem Scale, SRT/BP=Severe Restriction for Thinness/Binging and Purging, t=time point, TFEQ-R-18=Three-Factor-Eating-Questionnaire-Revised, WI=Whiteley Index to measure hypochondriacal worries and beliefs, YFAS=Yale Food Addiction Scale

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/