Measuring orthorexia nervosa: A comparison of four self-report questionnaires

Adrian Meulea,b,∗, Christina Holzapfelc, Beate Brandld, Martin Greetfeldb, Johannes Baltasar Hessler-Kaufmanna,b, Thomas Skurkd, Norbert Quadflieg, Sandra Schlegla, Hans Haunerc,e, Ulrich Voderholzera,b,f

a Department of Psychiatry and Psychotherapy, University Hospital, LMU Munich, Munich, Germany
b Schön Clinic Roseneck, Prien am Chiemsee, Germany
c Institute for Nutritional Medicine, School of Medicine, Technical University of Munich, Munich, Germany
d ZIEL – Institute for Food & Health, Technical University of Munich, Freising, Germany
e Chair of Nutritional Medicine, Else Kröner-Fresenius Center for Nutritional Medicine, Technical University of Munich, Freising-Weihenstephan, Germany
f Department of Psychiatry and Psychotherapy, University Hospital of Freiburg, Freiburg, Germany

ARTICLE INFO

Keywords:
Orthorexia
Self-report measures
Reliability
Validity
Psychometrics
Item analysis

ABSTRACT

Orthorexia nervosa is characterized by a preoccupation to eat healthily. However, reliability and validity of some of the existing measures of orthorexic symptomatology are questionable. Therefore, the aim of the current study was to examine internal reliability of and intercorrelations between four of the most popular self-report scales for measuring orthorexia nervosa: Bratman’s Orthorexia Test (BOT), the ORTO–15, the Eating Habits Questionnaire (EHQ), and the Düsseldorf Orthorexia Scale (DOS). Five-hundred and eleven adults (63% female) completed all four instruments. Model fit of the originally proposed factor structures of the BOT, DOS, and EHQ was good but was unacceptable for the ORTO–15. Similarly, internal reliability was good for the BOT, EHQ, and DOS, but was unacceptable for the ORTO–15. The BOT, EHQ, and DOS were highly correlated with each other while correlations with the ORTO–15 were of medium size. A subsequent exploratory item analysis suggested that the poor psychometric properties of the ORTO–15 are largely due to the originally proposed scoring procedure. In conclusion, the BOT, EHQ, and DOS are internally reliable instruments that seem to measure the same construct—orthorexic eating behavior. In line with previous suggestions, we conclude that the ORTO–15 cannot be recommended for the measurement of orthorexia nervosa, at least not when the originally proposed scoring procedure is used.

1. Introduction

The term orthorexia nervosa was coined by Bratman (1997) and was shortly after popularized by a book by Bratman and Knight (2000). Orthorexia nervosa refers to an obsessional preoccupation with eating healthy foods, focusing on concerns regarding the quality and composition of meals. This behavior may include consuming an unbalanced diet due to beliefs about food “purity”, rigid avoidance of foods that are believed to be unhealthy, feelings of guilt and worries after transgressions, or intolerance of others’ food beliefs. Extreme orthorexic behaviors may lead to impairment of physical health due to malnutrition and to impairment of social, academic, or vocational functioning due to obsessional thoughts and behaviors focusing on beliefs about healthy eating (Dunn & Bratman, 2016). While orthorexia is not recognized as a distinct eating disorder in current diagnostic classification systems, it has received increased research interest in recent years (Bratman, 2017; Missbach & Barthels, 2017).

Several self-report measures for the assessment of orthorexia nervosa have been developed. First, Bratman and Knight (2000) described a 10-item questionnaire, which would later be denoted as the Bratman Orthorexia Test (BOT). Further, Donini, Marsili, Graziani, Imbriale, and Cannella (2005) proposed a 15-item questionnaire entitled ORTO–15. However, research on orthorexia nervosa based on these two measures has been plagued by methodological issues and inconsistent findings (e.g., regarding the prevalence of orthorexia nervosa). These issues are primarily grounded in the fact that psychometric properties of the BOT are largely unknown and that psychometric properties of the ORTO–15 seem to be poor (Dunn & Bratman, 2016; Missbach, Dunn, & König, 2017).
More recently developed questionnaires include the 21-item Eating Habits Questionnaire (EHQ; Gleaves, Graham, & Ambwani, 2013) and the 10-item Düsseldorf Orthorexia Scale (DOS; Barthels, Meyer, & Pietrowsky, 2015). The DOS was originally developed in German language and has recently started to gain international recognition (Chard, Hilzendegen, Barthels, & Stroebele-Benschop, 2019; He, Ma, Barthels, & Fan, 2019). Another questionnaire—the Barcelona Orthorexia Scale—is currently under development (Bauer, Fusté, Andrés, & Saldaña, 2019). Considering the increasing number of orthorexia nervosa measures, it is an open question if these measures all assess the same construct and have comparable psychometric properties. Thus, the current study aimed to contribute to answering this question by employing all four of the existing self-report measures for the assessment of orthorexia nervosa and examining their factor structure, internal reliability and intercorrelations. Another aim was to explore possible reasons for the low internal reliability of the ORTO-15 that has been reported in previous studies (Missbach et al., 2017).

2. Methods

2.1. Participants

Five-hundred and eleven individuals (63.4% female, n = 324) participated in this study. Note that sample size is smaller for some variables because of missing data, but is larger than 500 for each variable reported in this article. Mean age was 43.4 years (SD = 18.1, Range 18–84) and mean body mass index was 25.2 kg/m² (SD = 4.7, Range 17.6–51.2). Nine participants (1.8%) had completed lower school education [German: Hauptschule], 37 (7.2%) had completed middle school education [German: Realschule], 67 (13.1%) had completed higher school education [German: Gymnasium], 134 (26.2%) had completed vocational training, and 262 (51.3%) had a university degree (data missing for 2 participants, 0.4%).

2.2. Questionnaires

2.2.1. BOT

The BOT (Bratman & Knight, 2000) has 10 items that usually have a dichotomous response format (e.g., 0 = no, 1 = yes). Thus, higher sum scores represent stronger orthorexic tendencies. In the current study, we used a German translation by Barthels and Pietrowsky (2012). Moreover, we changed the dichotomous response format to the four-point scale format of the DOS to streamline completion of the questionnaires and increase comparability of the different measures.

2.2.2. ORTO–15

The ORTO–15 (Donini et al., 2005) has 15 items with the following response options: 1 = always, 2 = often, 3 = sometimes, 4 = never. According to Donini et al. (2005), six items need to be recoded before calculating sum scores. Four of them are reversely coded (items #2, #5, #8, and #9). However, two items (#1 and #13) have a rather unusual recoding procedure: 2 = always, 4 = often, 3 = sometimes, 1 = never. Further, Donini et al. (2005) suggest calculating sum scores, for which lower scores represent stronger orthorexic tendencies. In the current study, the German version of the ORTO–15 was used (Missbach et al., 2015) and we applied the scoring procedure as described by Donini et al. (2005), but coded items such that higher sum scores represent stronger orthorexic tendencies to facilitate interpretation of results. Moreover, we changed the response labels to those of the DOS to streamline completion of the questionnaires and increase comparability of the different measures.

2.2.3. EHQ

The EHQ (Gleaves et al., 2013) has 21 items with response options ranging from 1 = false, not at all true to 4 = very true. There are no inverted items and higher sum scores represent stronger orthorexic tendencies. In the current study, we used a German version of the scale, which was independently translated by four researchers (first language German, fluent in English). These researchers then compared and discussed the four versions to achieve the most appropriate translation. Two native speakers (first language English, fluent in German) then counterchecked the final version. Moreover, we changed the response labels to those of the DOS to streamline completion of the questionnaires and increase comparability of the different measures.

2.2.4. DOS

The DOS (Barthels et al., 2015) has 10 items with the following response options: 1 = this does not apply to me, 2 = this does rather not apply to me, 3 = this does somewhat apply to me, 4 = this applies to me. There are no inverted items and higher sum scores represent stronger orthorexic tendencies. As the questionnaire was developed in German, we used the original version by Barthels et al. (2015).

2.3. Procedure

The study was approved by the institutional review board of the University of Munich (#17-544) and the Technical University of Munich (#492/17S). Participants were recruited from two studies on health and nutrition at the Technical University of Munich (cf. Brandl et al., submitted; Drabsch et al., 2018). The questionnaires were completed as paper-and-pencil versions either at the study center or were mailed to the participants. To those who received the questionnaires by mail, a reminder was sent if they had not responded after one month. All participants provided written informed consent. Participants completed some sociodemographic questions, followed by the ORTO–15, BOT, DOS, and EHQ. The survey also included further questions on nutrition and well-being, which are not reported here.

2.4. Data analyses

Statistical analyses were conducted with the program JASP version 0.10.2 (https://jasp-stats.org). Confirmatory factor analyses were run with JASP’s structural equation modeling module, which is based on the R-package lavaan (http://lavaan.ugent.be). Diagonally Weighted Least Squares was chosen as estimation method because of the ordinal scale structure (Li, 2016). A one-factor model was tested for the ORTO–15, BOT, and DOS, respectively. For the EHQ, a model was specified that included three first-order latent factors (representing the factors knowledge of healthy eating, problems associated with healthy eating, and feeling positively about healthy eating, as has been originally proposed; Gleaves et al., 2013) and a second-order factor (comprised of all three first-order factors). Model fit was evaluated according to the guidelines reported by Schermelleh-Engel, Moosbruger, and Müller (2003).

To evaluate internal reliability of all questionnaires, we report the greatest lower bound and McDonald’s ω in addition to Cronbach’s α, as has been recommended (Peters, 2014; Revelle & Zinbarg, 2008; Sijtsma, 2008). To examine intercorrelations between all questionnaires, we report Pearson’s correlation coefficients. Finally, we examined factor loadings of ORTO–15 raw scores (i.e., not the recoded items) using principal component analysis with extraction of one factor. Note that the aim of this analysis was to explore which items would show negative factor loadings and, thus, should be inversely scored. That is, the aim was not to explore factor structure of the ORTO–15 and determine how many factors should be extracted, which is why we restricted this analysis to extraction of one factor. As additional item statistics, we report corrected item-total correlations and Cronbach’s α if item deleted.
3. Results

3.1. Factor structure

For the one-factor structure of the ORTO–15, some indices suggested acceptable model fit (RMSEA = 0.06, GFI = 0.96) but others suggested unacceptable model fit (NFI = 0.79, NNFI = 0.82, CFI = 0.85). Importantly, standardized factor loadings ranged from −0.65 to 0.48, which further indicated the inappropriateness of a unidimensional structure (Fig. 1). The one-factor structure of the BOT had good model fit (RMSEA = 0.04, GFI = 0.99, NFI = 0.96, NNFI = 0.98, CFI = 0.99). All standardized factor loadings were higher than 0.27 (Fig. 2). Model fit was also good for the one-factor structure of the DOS (RMSEA = 0.03, GFI = 0.99, NFI = 0.97, NNFI = 0.99, CFI = 0.99). All standardized factor loadings were higher than 0.39 (Fig. 3). The model for the EHQ with three first-order factors and a second-order factor also fitted the data well (RMSEA = 0.04, GFI = 0.98, NFI = 0.96, NNFI = 0.98, CFI = 0.99). All standardized factor loadings were higher than 0.43 (Fig. 4).

3.2. Internal reliability of and correlations between measures

Internal reliability indices were acceptable or good for the BOT, EHQ, and DOS and unacceptable for the ORTO–15 (Table 1). Sum scores of the BOT, EHQ, and DOS were highly intercorrelated (rs > 0.70) while correlations with the ORTO–15 were of medium size (Table 1).

3.3. Exploratory item analysis of the ORTO–15

Principal component analysis revealed that two items (#5 and #8) had negative factor loadings, indicating that responses to these items should be inverted (Table 2). After recoding, internal reliability of the ORTO–15 items was acceptable (Cronbach’s α = 0.72, greatest lower bound = 0.82, McDonald’s ω = 0.72) and sum scores were highly correlated with the BOT (r = 0.75, p < .001), EHQ (r = 0.66, p < .001), and DOS (r = 0.71, p < .001). However, several items still had very low corrected item–total correlations (particularly items #8 and #14), indicating that not only the scoring procedure accounts for the scale’s poor psychometric properties.

4. Discussion

The current study investigated factor structure and internal reliability of as well as intercorrelations between four self-report measures for orthorexia nervosa. The BOT, EHQ, and DOS had a unidimensional structure (higher ordered for the EHQ), good internal reliability and sum scores were highly correlated with each other (rs > 0.70), indicating that they essentially measure the same construct (Vainik & Meule, 2018; Vainik, Neseliler, Konstabel, Fellows, & Dagher, 2015). In contrast, the ORTO–15’s factor structure had unacceptable model fit and the scale had low internal reliability and showed medium-sized correlations with the other three orthorexia nervosa questionnaires, suggesting that it measures a different construct and/or captures orthorexic tendencies less precisely.

The unreliability of the ORTO–15 found in the current study is in line with the majority of previous studies (Missbach et al., 2017; Valente et al., 2019). Yet, while previous studies tried to circumvent this issue by deletion of items (e.g., Alvarenga et al., 2012; Arusoğlu, Kabakci, Köksal, & Merdol, 2008; Missbach et al., 2015; Varga, Thege, Dukay-Szabó, Türy, & van Furth, 2014), it has not been scrutinized whether the poor internal reliability may be simply due to the scoring procedure. In line with the scoring procedure proposed by Donini et al. (2005), factor analysis of raw scores suggested that two items (#5 and #8) should be inversely scored. However, analyses also suggested that four other items that are usually recoded contribute to the scale’s unreliability. In fact, when recoding only items #5 and #8, the questionnaire performed much better in terms of internal reliability indices.
Fig. 3. Standardized factor loadings of the one-factor model for the DOS.

Fig. 4. Standardized factor loadings of model for the EHQ with three first-order latent factors and a second-order factor.

Table 1
Descriptive statistics, internal reliability, and correlations of four self-report measures for orthorexia nervosa.

<table>
<thead>
<tr>
<th>Measure</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
<th>Cronbach's α</th>
<th>Greatest lower bound</th>
<th>McDonald's ω</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bratman Orthorexia Test</td>
<td>509</td>
<td>18</td>
<td>5</td>
<td>10-35</td>
<td>.79</td>
<td>.81</td>
<td>.80</td>
<td>.47</td>
<td>.72</td>
<td>.78</td>
<td></td>
</tr>
<tr>
<td>ORTO–15</td>
<td>501</td>
<td>37</td>
<td>4</td>
<td>26-52</td>
<td>.33</td>
<td>.60</td>
<td>.44</td>
<td>.47</td>
<td>.51</td>
<td>.45</td>
<td></td>
</tr>
<tr>
<td>Eating Habits Questionnaire</td>
<td>511</td>
<td>38</td>
<td>10</td>
<td>21-70</td>
<td>.90</td>
<td>.94</td>
<td>.90</td>
<td>.72</td>
<td>.51</td>
<td>.79</td>
<td></td>
</tr>
<tr>
<td>Düsseldorf Orthorexia Scale</td>
<td>511</td>
<td>17</td>
<td>5</td>
<td>10-34</td>
<td>.84</td>
<td>.90</td>
<td>.86</td>
<td>.78</td>
<td>.45</td>
<td>.79</td>
<td></td>
</tr>
</tbody>
</table>

Notes. All correlation coefficients were statistically significant at p < .001.

* Note that—as opposed to the scoring procedure by Donini et al. (2005)—the ORTO–15 was coded such that higher values represent higher orthorexic tendencies.
Table 2

<table>
<thead>
<tr>
<th>Item</th>
<th>Raw scores</th>
<th>After recoding</th>
<th>Corrected item α</th>
<th>Cronbach's deleted total correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. When eating, do you pay attention to the calories of the food?</td>
<td>.53</td>
<td>.34</td>
<td>.67</td>
<td>.53</td>
</tr>
<tr>
<td>2. When you go in a food shop, do you feel confused?</td>
<td>.30</td>
<td>.22</td>
<td>.69</td>
<td>.30</td>
</tr>
<tr>
<td>3. In the last 3 months, did the thought of food worry you?</td>
<td>.59</td>
<td>.41</td>
<td>.66</td>
<td>.59</td>
</tr>
<tr>
<td>4. Are your eating choices conditioned by your worry about your health status?</td>
<td>.64</td>
<td>.45</td>
<td>.66</td>
<td>.64</td>
</tr>
<tr>
<td>5. Are you willing to spend more money to have healthier food?</td>
<td>.22</td>
<td>.11</td>
<td>.70</td>
<td>.22</td>
</tr>
<tr>
<td>6. Are you willing to have healthier food?</td>
<td>.60</td>
<td>.22</td>
<td>.70</td>
<td>.22</td>
</tr>
<tr>
<td>7. Does the thought about food worry you for more than 3 h a day?</td>
<td>.59</td>
<td>.43</td>
<td>.63</td>
<td>.59</td>
</tr>
<tr>
<td>8. Do you allow yourself any eating transgressions?</td>
<td>.58</td>
<td>.24</td>
<td>.69</td>
<td>.58</td>
</tr>
<tr>
<td>9. Do you think your mood affects your eating behavior?</td>
<td>.46</td>
<td>.41</td>
<td>.66</td>
<td>.46</td>
</tr>
<tr>
<td>10. Do you think that consuming healthy food may improve your appearance?</td>
<td>.66</td>
<td>.45</td>
<td>.66</td>
<td>.66</td>
</tr>
<tr>
<td>11. Do you think that the conviction to eat only healthy food increases self-esteem?</td>
<td>.44</td>
<td>.28</td>
<td>.69</td>
<td>.44</td>
</tr>
<tr>
<td>12. Do you think that consuming healthy food may improve your appearance?</td>
<td>.58</td>
<td>.24</td>
<td>.69</td>
<td>.58</td>
</tr>
<tr>
<td>13. Do you feel guilty when transgressing?</td>
<td>.55</td>
<td>.22</td>
<td>.69</td>
<td>.55</td>
</tr>
<tr>
<td>14. Do you think that on the whole there is also unhealthy food?</td>
<td>.52</td>
<td>.19</td>
<td>.69</td>
<td>.52</td>
</tr>
<tr>
<td>15. At present, are you alone when having meals?</td>
<td>.52</td>
<td>.19</td>
<td>.69</td>
<td>.52</td>
</tr>
</tbody>
</table>

and correlations with other measures for orthorexia nervosa than when using the originally proposed scoring procedure. Yet, some items still had very low corrected item–total correlations, suggesting that there are other issues beyond the scoring procedure that make the ORTO–15 an unsound measure. These issues may include unclear item wordings or items that are not generic for the orthorexia construct.

Interpretation of our findings is limited by several factors. First, we used German versions of all questionnaires and, thus, results may be different with other language versions. Second, previous studies have usually employed a dichotomous response format for the BOT and, thus, we cannot exclude the possibility that the sound psychometric properties of the BOT in the current study may be due to the four-point scale response format. Similarly, we slightly modified the response labels of the ORTO–15 and the EHQ—yet maintained the four-point scale format—which might have influenced results as well. However, the current results are in line with the large majority of studies on psychometric properties of the ORTO–15 and, in addition, the replicated factor structures of and high correlations between the BOT, DOS, and EHQ indicate both factorial and convergent validity of the versions used in the current study. Third, while we investigated a very broad sample (e.g., regarding age and body mass index), it was not nationally representative and tended to be rather highly educated. Thus, it may be that psychometric properties of self-report measures for orthorexia nervosa may be different in less educated samples. Finally, we did not examine indicators of validity that might differ between the four questionnaires. This is certainly a challenge for research on orthorexia nervosa in general, as it is for any field where new concepts of mental disorders are proposed (e.g., behavioral addictions; Billieux, Schimmenti, Khazaal, Maurage, & Heeren, 2015). Thus, a future direction would be to develop other assessment methods for orthorexic behavior (e.g., interview techniques) that might be used for testing predictive or criterion validity of the self-report instruments.

In conclusion, we found that the BOT, EHQ, and DOS are internally reliable self-report instruments and their high intercorrelations support convergent validity. In contrast, the ORTO–15 is internally unreliable, which is primarily due to the way items are usually scored. Thus, research on orthorexia nervosa should prefer using the BOT, EHQ, or DOS.

Acknowledgments

Studies were funded in part by a grant of the German Ministry for Education and Research (BMBF, 01EA1409A), by the Else Kröner-Fresenius-Foundation (Bad Homburg, Germany), and by the Helmholtz cross-program topic Metabolic Dysfunction. Data related to this paper were in part gathered within the framework of the enable cluster (http://enable-cluster.de). The authors thank Margot Maier and Theresa Drabsch for supporting data collection.

References


A. Mesle, et al.  
Appetite 146 (2020) 104512
4.2015.009.


Eating and Weight Disorders, 22. https://doi.org/10.1007/s40519-017-0365-1 1–1.


