Empirical investigation of different factor structures for the eating disorder examination–questionnaire in adult women with anorexia nervosa

Article

Accepted Version

Jenkins, P. E. ORCID: https://orcid.org/0000-0003-1673-2903, Blalock, D. V., Duffy, A., Mehler, P. S. and Rienecke, R. D. ORCID: https://orcid.org/0000-0003-0423-7403 (2023) Empirical investigation of different factor structures for the eating disorder examination–questionnaire in adult women with anorexia nervosa. Assessment. ISSN 1552-3489 doi: 10.1177/10731911231198207 Available at https://centaur.reading.ac.uk/113420/

Empirical Investigation of Different Factor Structures for the Eating Disorder Examination Questionnaire (EDE-Q) in Adult Women with Anorexia Nervosa

Paul E. Jenkins[1], Dan V. Blalock[2,3], Alan Duffy[4], Philip S. Mehler[4,5,6], & Renee D. Rienecke[4,7]

[1]School of Psychology and Clinical Language Sciences, University of Reading, Reading. RG6 6ES. United Kingdom
[2]Center of Innovation to Accelerate Discovery and Practice Transformation, Durham Veterans Affairs Medical Center, Durham, NC, USA
[3] Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA
[4]Eating Recovery Center/Pathlight Mood & Anxiety Center, USA
[5]ACUTE Center for Eating Disorders at Denver Health, Denver, CO, USA
[6]University of Colorado School of Medicine, Denver, CO, USA
[7]Department of Psychiatry and Behavioral Sciences, Northwestern University, Chicago, IL, USA

Corresponding Author:
Renee D. Rienecke
333 N. Michigan Avenue, Ste. 1900, Chicago IL 60601
Renee.Rienecke@ercpathlight.com
312-964-4644

Author Note:
Deidentified data are available upon reasonable request by emailing the corresponding author.

Funding: Dr. Blalock was supported by Career Development Award 19–035 (IK2HX003085-01A2) from the United States Department of Veterans Affairs Health Services Research and Development (HSR&D) Service.

Abstract

The Eating Disorder Examination Questionnaire (EDE-Q) is a widely used self-report measure of eating pathology. Despite widespread use, investigations of its factor structure have proved inconclusive and rarely supported the 'original' interpretation. The current study evaluates several proposed factor solutions of the EDE-Q using latent variable analysis in a sample of adult women with anorexia nervosa (AN). 804 patients from a specialist treatment center in the USA participated in the study. Confirmatory factor analysis was conducted on 22 EDE-Q items assessing attitudinal features of eating pathology. Findings suggested that three full-item versions (none of which was the 'original' interpretation) fit the data adequately, with a brief, seven-item version showing excellent fit. The study is one of the first to examine this within a sample of women with AN and provides an empirical foundation for how best to use the EDE-Q among clinical and research participants with AN. Findings suggest that the 'original' factor structure lacks structural validity in women with AN. Its use should generally be discouraged, and future work on screening and treatment outcomes might consider the EDE-Q7.

Keywords: eating disorders, confirmatory factor analysis, Eating Disorder Examination–Questionnaire, psychometrics

The term eating disorders (EDs) refers to a collection of psychiatric illnesses with a combined lifetime prevalence of nearly 15% in young women (Stice et al., 2013). Anorexia nervosa (AN) is an ED characterised by disturbed eating behaviour (e.g., persistent restriction of food intake) that effects low body weight (or interferes with expected weight gain) in the presence of ED cognitions, such as a fear of becoming fat (American Psychiatric Association [APA], 2013). Psychiatric and physical comorbidity is common (e.g., see Meczekalski et al., 2013). The assessment of EDs is complex and especially so for AN in light of the symptoms commonly reported. For example, whilst several self-report measures have been designed to assess the profile and extent of eating pathology, individuals with AN may "minimize" symptoms and be less likely to endorse certain items (Passi et al., 2003), compromising the reliability of such assessments.

One of the most widely used self-report measures is the Eating Disorder Examination Questionnaire (EDE-Q; Fairburn & Beglin, 1994; currently version 6.0; Fairburn & Beglin, 2008), which has been found to perform similarly to clinician-administered interviews, particularly in the assessment of behavioral symptoms (Wolk et al., 2005). The EDE-Q consists of 22 items addressing attitudinal symptoms, scored on a 7-point forced-choice scale (Fairburn & Beglin, 1994), contributing to four subscales: (Dietary) Restraint; Eating Concern; Shape Concern; and Weight Concern (Fairburn & Beglin, 2008). Six additional questions address behavioral features of an ED ('disordered eating behaviors'), such as binge eating; these items are scored on a ratio scale and are typically not included in calculation of subscales, although their addition is recommended for clinical purposes (e.g., Kliem et al., 2016). A Global score – "indicative of overall eating problems" (Friborg et al., 2013, p. 199) – can be calculated as a mean of the four subscales (Fairburn & Beglin, 2008), although lack of support in factor analysis studies has called into question whether eating pathology can be

accurately summarised according to these suggestions (Jenkins & Rienecke, 2022; Rand-Giovannetti et al., 2020).

Psychometric studies of the EDE-Q (see Berg et al., 2012) have generally offered support for the convergent reliability, discriminant validity (e.g., whether the EDE-Q is empirically distinct from another scale), temporal stability, and internal consistency of the subscales. Despite these encouraging findings, investigations of the underlying factor structure (structural validity) of the EDE-Q have been less convincing (see Jenkins & Rienecke, 2022, for a review). The 'original' factor structure has received only limited empirical support and few of the novel models suggested (typically through exploratory factor analysis; EFA) have been replicated in separate samples (Carey et al., 2019; Rand-Giovannetti et al., 2020), suggesting a lack of consistency and difficulty drawing "firm conclusions regarding an optimal factor solution" (Allen et al., 2011, p. 144). Briefer versions, comprising a subset of EDE-Q items, have, however, performed more consistently well and a seven-item version (the EDE-Q7; Grilo et al., 2013) offers particular promise for capturing treatment-related changes and screening for EDs (Machado et al., 2020). Compared to the Original model, the EDE-Q7 has demonstrated more variable patterns of correlations with related constructs, such as depression, which suggests "better convergent and discriminant validity" of the briefer version (Grilo et al., 2015, p. 287). Moreover, whilst few studies have explicitly evaluated the effects of translation, the EDE-Q7 has been shown to demonstrate language invariance in a sample of Hispanic/Latino/a/x persons seeking bariatric surgery in the USA (Marek et al., 2023), suggesting further reliability of its structure.

In addition to inconsistent findings regarding the subscales, presence of a 'global' latent variable has not been reliably estimated from factor analysis studies. When higher-order factor structures of the EDE-Q have been tested (i.e., with subscales loading onto a

'Global' scale), results have indicated worse fit than first-order models (e.g., Rand-Giovannetti et al., 2020). A bifactor model (where four factors exist alongside an uncorrelated general latent factor) demonstrated good fit in a community sample of 3000 Norwegian women (Friborg et al., 2013; see also Nagata et al., 2023).

In accord with studies of the factor structure of the EDE-Q more generally, those including mixed clinical samples (i.e., without stratifying by diagnosis) have produced inconsistent results. Calugi et al. (2017), for example, investigated the factor structure of the EDE-Q using confirmatory factor analysis (CFA) in a mixed inpatient and outpatient sample, around three-quarters of whom met criteria for AN. Comparison of three models (the Original model, the one-factor solution, and the brief EDE-Q7 mentioned above) found that only the brief version demonstrated adequate fit. One possible benefit of the EDE-Q7 is that it could more accurately assess core symptoms of AN, given that the appropriateness of other measures has been questioned (e.g., Surgenor & Maguire, 2013). Existing findings in clinical samples also suggest that concern with shape or weight (often believed to be overlapping constructs; e.g., Parker et al., 2016) is distinct from dissatisfaction with one's appearance (e.g., Lydecker et al., 2017), as reflected in the interpretation of the EDE-Q7. Given that shape/weight overvaluation is considered an indicator of eating pathology (APA, 2013; Hrabosky et al., 2008), this distinction, if supported, may hold promise for identifying reliable predictors of treatment outcome.

Although a number of studies have evaluated the factor structure of the EDE-Q with mixed or non-clinical samples, relatively few have focused on individuals with AN. In a sample of 169 adult female inpatients, Phillips et al. (2018) generated a novel four-factor solution using EFA. Whilst two factors showed similarities to Restraint and Eating Concern as outlined by Fairburn and Beglin (2008), a third seemed to represent dissatisfaction with body image (labelled 'Appearance Concern' by other authors; e.g., Hrabosky et al., 2008;

Parker et al., 2016), and a fourth combined several items from the Weight Concern and Shape Concern subscales (see also Peterson et al., 2007). A Dutch study using a sample of female patients with AN (Pennings & Wojciechowski, 2004) suggested the presence of a 'Global' score by identifying a one-factor solution of the EDE-Q.

Studies have investigated the EDE-Q7 among bariatric surgery patients, mixed clinical samples, community samples, undergraduate students, sexual minority men and women, and different dietary groups (e.g., vegans, omnivores). To the authors' knowledge, the EDE-Q7 has not been comprehensively evaluated in an exclusively AN sample, and so further investigation in such individuals seems warranted. The findings could inform how best to use self-report measures (particularly variants of the EDE-Q) in clinical settings, improve screening, and accurately measure outcome following treatment.

A recent review (Jenkins & Rienecke, 2022) highlighted several distinct models of the EDE-Q that have been proposed in the literature and so, in the current study, the most consistently supported factor structures will be examined:

- Model 1 ('Original'): The 'Original' model of Fairburn and Beglin (2008), including all 22 attitudinal items and comprising four latent variables, as detailed above (see also Table 1);
- Model 2 ('Friborg'): A four-factor model (22 items) described by Friborg et al. (2013), with some differences to the Original, such as a 'combined' Weight and Shape Concern scale;
- Model 3 ('Peterson'): Peterson et al.'s (2007) three-factor model (22 items), with most Shape Concern and Weight Concern items loading on one factor;
- Model 4 ('EDE-Q-7'): Grilo et al.'s (2013) seven-item version, comprising three factors labelled Dietary Restraint, Body Dissatisfaction, and Shape/Weight Overvaluation.

In addition, two 22-item models with less evidence were included given that they were developed from samples of women with AN: Model 5 ('Phillips'), a four-factor model proposed by Phillips et al. (2018); and Model 6 ('One-factor'), a one-factor model, proposed by Pennings and Wojciechowski (2004). Model compositions, along with item labels based on the EDE-Q (Fairburn & Beglin, 2008), are shown in Table 1.

We also aim to investigate the fit of bifactor solutions for the EDE-Q7 (Model 7: 'EDE-Q7-Bifactor') and Friborg models (Model 8: 'Friborg-Bifactor'), given the suggested theoretical interpretations and previous work (e.g., see Friborg et al., 2013). As interpretations of the EDE-Q often suggest that a 'Global' score can be obtained by summing the means of the subscale scores and dividing this by four (the number of subscales) (Fairburn & Beglin, 2008), we will test higher-order solutions of models showing acceptable fit in the first-order analyses and compare these to their 'first-order' alternative (e.g., see Rand-Giovannetti et al., 2020). To explore validity of empirically supported versions, we will also investigate correlations between different constructs, such as depression and disordered eating behaviours, and alternative versions of the EDE-Q. Given previous work, we expected to find significant and strong correlations with depression symptoms (e.g., Machado et al., 2020) and no significant correlations with age (e.g., Jenkins & Davey, 2020). Whilst some studies have shown significant correlations with body mass index (BMI) within ED samples (e.g., Grilo et al., 2013), this may not be evident in individuals presenting within a constrained weight range (i.e., underweight) and, similarly, behaviours such as binge eating might also be infrequent in an AN sample (e.g., Binford et al., 2005). Finally, in order to ascertain whether the EDE-Q can be interpreted as equivalent across samples, we planned to test whether factor structure(s) showing acceptable fit were invariant across the restricting and binge-purge subtypes of AN.

Given that all the suggested models have a degree of empirical support for structural validity (with the exception of the Original model) across a range of samples, it is difficult to offer clear hypotheses. Nonetheless, we expect to find positive correlations between the EDE-Q (sub)scales and measures of disordered eating behaviors and comorbid psychopathology. It is also anticipated that the Original model will show poor fit and the EDE-Q7 will show good fit (e.g., Calugi et al., 2017). Those models combining items from the Weight and Shape Concern subscales (e.g., Friborg et al., 2013; Peterson et al., 2007) may also perform well (see Jenkins & Rienecke, 2022).

Method

Participants

Participants were 915 adult women seeking treatment at Eating Recovery Center (ERC), a large multisite treatment facility offering higher levels of care for patients with EDs, between December 2019 and September 2022. Consecutive referrals to ERC, aged 18 years or older, completed baseline measures as part of the admission process within five days of admission. Patients who signed research consent, had a diagnosis of AN, and completed the EDE-Q were included in the study. Diagnoses were made by their treating clinician according to DSM-5 (APA, 2013) criteria. This study was approved by Salus Institutional Review Board and was not preregistered.

Of these 915, 111 were missing sufficient data on the EDE-Q for analysis, with 108 (11.8%) not completing the questionnaire. Of the 804 used in the final analysis, six individuals (0.75%) received a diagnosis of AN, with a further 539 (67.04%) the restricting subtype (AN-R) and 259 (32.21%) the binge-purge subtype (AN-BP). 780 (97.0%) identified with the female gender, four (0.5%) identified as male or transgender (female-to-male), and 20 (2.5%) as non-binary. The majority (670; 83.3%) identified as White, 38

(4.7%) as Hispanic or Latino, 18 (2.2%) as Biracial/Multiracial/Mixed Race, 17 (2.1%) as Asian, 10 (1.2%) as Black or African American, one (0.1%) as American Indian/Native American or Alaska Native, and one (0.1%) as Native Hawaiian/Pacific Islander. Racial / ethnicity data on 45 individuals (5.6%) were missing. Age of the sample ranged from 18 to 64 years (mean = 25.65; SD = 9.51; median = 22.00). Mean BMI (kg/m^2) on admission for the overall sample (N = 750) was 16.56 (SD = 1.57, range = 11.2–19.0, median = 16.75); this was calculated from admission height and weight, taken by clinical staff.

Measures

The EDE-Q (Fairburn & Beglin, 1994, 2008) is a widely used self-report questionnaire assessing behavioral and attitudinal features of eating pathology. As mentioned above, it has been shown to demonstrate good psychometric properties (Berg et al., 2012). In Version 6.0 (Fairburn & Beglin, 2008), used in the current study, twenty-two items assess attitudinal features and a further six items assess disordered eating behaviors, a measurement distinction made in the development of the EDE-Q (Fairburn & Beglin, 1994).

As part of the examination of validity, the Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001) was included, using the Total score as a continuous variable. The PHQ-9 is a self-report measure of depression severity, with nine items measured on a 0 (not at all) to 3 (nearly every day) scale and assessing symptoms such as “feeling tired or having little energy” over the previous two weeks. Higher scores indicate greater depressive symptomatology. Depression categories include “minimal depression” (scores of 1-4), “mild depression” (scores 5-9), “moderate depression” (scores 10-14), “moderately severe depression” (scores 15-19), and “severe depression” (scores 20-27). The PHQ-9 has been shown to have excellent internal reliability and test-retest reliability (Kroenke et al., 2001).

Statistical analyses

All factor structures were derived from the 22 attitudinal items of the EDE-Q and CFA was conducted to evaluate the fit of the models described above (see also Table 1) using the R *lavaan* package (version 0.6-9; Rosseel, 2012). Packages *psych* (Revelle, 2021), *MVN* (4.0.5; Korkmaz et al., 2014), and *correlation* (Makowski et al., 2019) were used to assess internal consistency and the Kaiser-Meyer-Olkin (KMO) statistic, multivariate normality, and correlations, respectively. Mardia's test suggested the presence of non-normality, with significant skew and kurtosis, so robust estimation (Weighted Least Squares Means- and Variance-adjusted; WLSMV) was used in CFA. To identify the model, the residual variances of each latent variable were fixed to 1 and allowed to correlate (e.g., Bollen, 1989) and endogenous variables were treated as categorical (using the 'Ordered' function in *lavaan* [Rosseel, 2012]). A sample of at least 153 is required to provide 95% power for a Root Mean Square Error of Approximation (RMSEA) estimate of 0.05 with an α of 0.05 (Moshagen & Erdfelder, 2016).

Model fit was assessed using the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Standardised Root Mean Squared Residuals (SRMR) and RMSEA (including 90% confidence intervals) statistics, in line with common practice and recommendations for examining model fit (Brown, 2015; Hu & Bentler, 1999). The χ^2 value is reported but, due to concerns with large sample sizes, some caution is warranted (e.g., see Shi et al., 2018). Conventional cutoffs (Hu & Bentler, 1999) are used to inform decisions regarding model fit, with subsequent interpretations based on both statistical and theoretical standpoints (e.g., Schmitt et al., 2018). We did not consider modification indices as the study aimed to investigate the performance of several interpretations of the EDE-Q, rather than identify an optimal factor structure based on (potentially minor) modifications. However, we did aim to

examine the difference in fit between higher-order models and their 'first-order' alternative, and report both the χ^2 and Satorra-Bentler χ^2 (Satorra & Bentler, 2001) difference tests. As the scoring criteria for the EDE-Q (Fairburn & Beglin, 2008) suggest that a second-order structure might be imposed on EDE-Q data, we aimed to investigate higher-order solutions for Model 1 ('Original'), Model 2 ('Friborg'), and Model 3 ('Phillips') models if acceptable fit is demonstrated. It is not possible to test higher-order models of the EDE-Q7 or Peterson et al. solutions: as both have three first-order factors, the higher-order solution has the same degrees of freedom (due to an identical number of covariances among higher-order constructs) and will give the same fit statistics (see also Brown, 2015).

The omega coefficient was used as an estimate of internal consistency reliability for each model, using the *MBESS* package (Kelley, 2017), with $\omega_{u\text{-}cat}$ for first-order models and Omega-higher-order (ω_{ho}) for higher-order models (Flora, 2020). Multigroup analyses were planned with those model(s) showing acceptable fit to test measurement invariance across AN diagnoses (AN-R versus AN-BP). Given that a subset of participants did not identify with the female gender, we planned to repeat assessments of model fit (i.e., confirmatory factor analyses) with cisgender individuals. Correlations among the sum scores of acceptable model(s) were assessed using non-parametric (Spearman's ρ) tests with age, BMI, depression, and disordered eating behaviors (derived from the behavioral items of the EDE-Q). It was expected that the EDE-Q would show significant correlations with depression, which would be comparable among adequately-fitting models (e.g., Grilo et al., 2015; Machado et al., 2020).

Results

Confirmatory factor analysis

The overall KMO statistic for EDE-Q items was 0.95 (range = 0.91 – 0.99), suggesting that the data were appropriate for factor analysis. Findings relating to each model tested were as follows:

- *Model 1*. The factor structure for the Original model was nonpositive definite[1], which renders the model non-admissible and is suggestive of 'over-factoring' (e.g., see Brown, 2015). Specifically, there was a strong correlation between Weight Concern and Shape Concern (r = 1.0) and high factor loadings. Given that a model combining the Weight Concern and Shape Concern subscales (Peterson et al., 2007) was also being tested, we did not perform any modifications to the Original model, nor did we test a higher-order alternative, and fit statistics are not reported (e.g., see Brown, 2015).
- *Model 2*. The four-factor model of Friborg et al. (2013) showed adequate fit (see cutoffs provided in Table 2, based on Hu & Bentler, 1999).
- *Model 3*. The three-factor model of Peterson et al. (2007) also showed adequate fit.
- *Model 4*. Fit for the brief, seven-item version (Grilo et al., 2013) of the EDE-Q was deemed excellent.
- *Model 5*. The model of Phillips et al. (2018) also showed adequate fit.
- *Model 6*. The one-factor model of Pennings and Wojciechowski (2004) demonstrated poor fit, with all indices falling below recommended cutoffs.
- *Model 7*. The EDE-Q7 with a bifactor solution did not converge.
- *Model 8*. The four-factor model of Friborg with a bifactor solution did not converge.

[1] For a formal definition and fuller explanation of this term, see Wothke (1993)

- *Model 9*. Given the acceptable fit of Model 2, a higher-order model of the solution proposed by Friborg et al. was tested and showed significantly worse fit compared to the first-order model ($\chi^2(2) = 61.23$, $p < 0.001$; $SB\chi^2(2) = 37.19$, $p < 0.001$),
- *Model 10*. A higher-order model based on that of Phillips et al. showed a similar pattern, fitting significantly worse compared to the first-order model (Model 5) ($\chi^2(2) = 35.61$, $p < 0.001$; $SB\chi^2(2) = 18.11$, $p < 0.001$).

In brief, the study found support for the 22-item models of Friborg et al. (2013; Model 2), Peterson et al. (2007; Model 3), and Phillips et al. (2018; Model 5); higher-order models (where the same first-order factors load onto a single [higher-order] factor) showed significantly worse fit than the corresponding first-order model. A brief version of the EDE-Q (Grilo et al., 2013; Model 4) showed excellent fit. We repeated the analyses with only those individuals identifying with the female gender, finding almost identical results (see Table S1).

Internal consistency reliability

As shown in Table 1, internal consistency estimates were acceptable for most latent variables. Inter-item correlations are presented in Figure S1. For the three 'full-item' models (Models 2, 3, and 5), estimates for one subscale (containing similar items across models) were typically lower than others. In the models of Friborg et al. (2013; Model 2) and Phillips et al. (2018; Model 5), one subscale was comprised of the same items (Items 19, 20, and 21), giving an omega value of 0.71. This scale includes three of the five items of the original Eating Concern subscale (Fairburn & Beglin, 2008). Similarly, omega was 0.78 for the seven-item scale described by Peterson et al. (2007; Model 3), which contained all five items of the original Eating Concern subscale.

The internal consistency estimate was high for the one-factor model (Pennings & Wojciechowski, 2004; Model 6), but this can be misleading when there is multidimensionality (Flora, 2020), which seems likely given the various supported interpretations of the EDE-Q. The internal consistencies of the 'global' factor of the higher-order models were fairly high (see Table 2).

Measurement invariance

Multigroup CFA was conducted for the diagnostic subgroups of AN-R and AN-BP (those with a diagnosis only of AN [n = 6] were excluded) in order to further test the psychometric properties of the EDE-Q. However, the models tested (Models 2 – 5) failed to converge, in part as no-one in the AN-BP group scored "1" on Q23 (meaning that the parameters across groups were different)[2].

Validity analyses

The sum scores of the EDE-Q7, Peterson, Phillips, and Friborg models (i.e., those showing acceptable fit) showed a similar pattern of correlations with demographic and symptom variables, suggesting similarity in their respective validity. All were significantly correlated with purging (both laxative use and vomiting), excessive exercise, and depression, with significant, but less strong, correlations with binge eating. There were inconsistent correlations with BMI and no significant correlations with age (see Table 3).

Discussion

Despite widespread evaluation of the psychometric properties of the EDE-Q (Fairburn & Beglin, 1994; 2008), few studies have investigated its factor structure – an indicator of structural validity – in clinical samples. The current study compared several proposed interpretations of the EDE-Q in a sample of women with AN, a notably under-studied group.

[2] It is possible to 'collapse' scoring groups although this was decided against due to potential loss of data

The study found support for three different full-item versions of the EDE-Q (Friborg et al., 2013; Peterson et al., 2007; Phillips et al., 2018) in a sample of women with AN, suggesting that variations on the factor structure of the 22 attitudinal items might not confer significant benefit in terms of statistical fit or practical utility (see also Figure S2). In addition to findings regarding structural validity, these versions showed similar patterns of correlations with demographic and symptom measures. Researchers should carefully consider whether generating a novel factor structure (through EFA, for instance) is truly indicated given the similarities of these statistically acceptable interpretations.

The 'Original' model did not converge, likely due to high factor loadings and correlations between latent variables, which are common findings in the literature (Allen et al., 2011; Parker et al., 2016). Thus, alternative versions appear to show more reliable structural validity and the findings cast further doubt on calculation of subscale scores according to the 'original' scheme (Friborg et al., 2013).

The full-item models of Friborg et al. (2013) and Phillips et al. (2018) endorsed a subscale comprised of the same three items, suggesting a degree of reliability. Whilst the relatively low internal consistency could be due to the small number of items, it could also suggest that the items were incorrectly specified on the latent factor or that this subscale (similar to 'Eating Concern'; Fairburn & Beglin, 2008; see also Peterson et al., 2007) may have less relevance for women with AN (e.g., items regarding eating in secret, or concern about others seeing the individual eat; see also Laskowski et al., 2023). Given that the fit of the three full-item models was similar, the current study offers further support for the reliability of Friborg et al.'s model, which consists of four factors characterised by the authors as Dietary Restraint, Eating Concern, Shape and Weight Concern, and "a mixture of five items from the [Eating Concern] and [Dietary Restraint] subscales, in addition to the preoccupation with weight/shape item" (p. 200), labelled by Rand-Giovannetti et al. (2020)

as Preoccupation and Restriction[3]. The current study is also (to the authors' knowledge) the first use of CFA with the model of Phillips et al. and suggests that this interpretation shows structural validity in a sample of adult women with AN.

A brief, seven-item version (EDE-Q7; Grilo et al., 2013) showed excellent fit, in line with existing studies across a range of samples (see Jenkins & Rienecke, 2022). In a large study of German women with EDs, Laskowski et al. (2023) concluded that the concepts of (dietary) restraint, shape and weight overvaluation, and body dissatisfaction appear quite stable across diagnostic groups. Interestingly, all are constructs uniquely assessed by the EDE-Q7, suggesting that this measure might reliably capture core eating psychopathology in a brief way and offer useful advantages to guide treatment monitoring (see also Machado et al., 2020). This theoretical structure also mirrors the 'transdiagnostic' approach to EDs whereby a 'core psychopathology' of overvaluation of eating, shape, weight, or their control is a primary maintenance factor, with other clinical features (including dietary restraint, for instance) driven by these standards (Fairburn et al., 2003).

In contrast, items relating to this core psychopathology often coalesce in latent factors of full-item models, and thus suggest some redundancy and also discontinuity between current theoretical foundations of eating pathology and what dimensions are assessed in practice. In addition to the potential for conflating distinct concepts (e.g., concerns about weight or shape), item redundancy could explain the under-reporting often seen in AN (if items are thought not to apply to an individual, for instance, or do not accurately represent ED pathology). Careful identification of relevant items would assist in both the refinement of measurement models and accurate assessment of eating pathology. In light of weak correlations between most latent variables and binge eating, it is apparent that measures of

[3] Rand-Giovannetti et al. (2020) and other authors (e.g., Klimek et al., 2021) have labelled the 'Eating Concern' factor in Friborg et al. 'Eating Shame' given that it is not identical to the Original subscale.

ED behaviours, such as binge eating, offer necessary information over and above the attitudinal items. Thus, it is important to investigate the performance of these measures in different diagnostic groups, as the EDE-Q7 (in the absence of other symptom measures) may not perform as well when utilised with individuals with binge-eating spectrum disorders, for instance, given inconsistent correlations with binge eating and strong correlations with purging.

Further support was found for the distinction between overvaluation of weight / shape (a core diagnostic construct across EDs; APA, 2013) and body dissatisfaction (a feature of the illness that commonly occurs; e.g., see Cooper & Fairburn, 1993; Mond & Hay, 2011). The two scales were correlated to a moderately strong degree ($\rho = 0.67$), supporting previous findings regarding the interpretability of the EDE-Q7 and the partial overlap between these constructs (e.g., Grilo et al., 2015; Machado et al., 2020; see also Mond & Hay, 2011). Thus, adoption of the EDE-Q7 might, in spite of its brevity, represent a useful means of predicting and evaluating treatment outcome (e.g., Machado et al., 2020). Of note, correlations were conducted with sum scores of the EDE-Q, as opposed to scores derived based on factor loadings (i.e., implying that "each item contributes an equal amount of information to the construct being measured" and thus being constrained; McNeish & Wolf, 2020, p. 2289; see also Widaman & Revelle, 2023).

The current study also evaluated models where a general factor ('Global') models the shared variance between subordinate ('first-order') factors. Whilst acceptable fit was demonstrated for both the 'Phillips' and 'Friborg' models, findings were in line with those of Rand-Giovannetti et al. (2020) who found that the higher-order model of Friborg et al. (2013) showed worse fit then the corresponding first-order model. As decisions around interpretation should not be made solely on model fit statistics (e.g., Markon, 2019), findings of the current study (such as the poorer fit of higher-order models and the generally

acceptable internal consistency reliability of the Global score) are in line with the view that the Global score of the EDE-Q might reflect an indicator of overall eating pathology (see Friborg et al., 2013), which is suitable for assessing treatment outcome, for instance. However, it might be more consistent with psychometric findings to calculate the Global score based on an average of all EDE-Q items rather than the subscales, particularly if calculating using the 'Original' model (e.g., Aardoom et al., 2012), which seems to be contraindicated. Whether this 'g' factor is validly calculated from briefer versions remains to be seen (and perhaps addressed with different methods, such as Item Response Theory; e.g., see O'Connor et al., 2022), particularly as the current study could not assess a bifactor model of the EDE-Q7.

Turning to clinical uses, whilst the predictive utility of brief versions of the EDE-Q remains to be fully demonstrated, particularly in AN, the Overvaluation of Weight and Shape subscale of the EDE-Q7 has been found to have diagnostic utility in both atypical eating disorders (Ortiz et al., 2021) and disorder severity in binge-eating disorder (Forrest et al., 2022). Short forms can also discriminate between clinical and non-clinical samples and are sensitive to change over treatment (Machado et al., 2020), suggesting that they might be reliable and valid measures of eating psychopathology. Previous authors have highlighted the risks associated with short versions (e.g., Goel et al., 2022), such as the reliability of subscales with only two items, so more thorough investigation of their clinical utility seems warranted.

Whilst the current study identified several versions of the EDE-Q that demonstrate structural validity in a sample of women with AN, longer versions may retain some item-level problems (Rand-Giovannetti et al., 2020). In the current study, means for Items 2 ('Have you gone for long periods of time [8 waking hours or more] without eating anything at all to influence your shape or weight?') and 19 ('Over the past 28 days, on how many days

have you eaten in secret [i.e., furtively]?') were particularly low. Further, 75 individuals (9.3%) reported a Global score <1 (equating to below the 35th percentile; e.g., Carey et al., 2019; Luce et al., 2008; see also Table S2), a proportion which was lower with the EDE-Q7 total score (n = 51; 6.3%). This suggests that a significant proportion of women with AN might score significantly below expected thresholds, perhaps due to symptom 'denial' or a reluctance to disclose (Surgenor & Maguire, 2013). Whilst cutoffs are often used to screen for and identify EDs, as well as indicating remission, suggested values using the EDE-Q have varied and may be lower for those with AN (e.g., see, Rø et al., 2015). The relatively high numbers of women with AN scoring below 4, for example, reflect concerns that have been highlighted with the use of cutoffs with the EDE-Q (Meule, 2021) and may warrant further scrutiny.

Several limitations should be noted. First, the current study could not determine the precedence of a bifactor model over a first-order or alternative second-order model. Whilst there are some conceptual and technical reasons to investigate bifactor models, such as interpretation of the 'Global' factor as correlated or uncorrelated to subscales (Friborg et al., 2013), any such study needs to consider relevant theory as well as the limitations and caveats of fitting bifactor models (e.g., Markon, 2019; Eid et al., 2017). Several additional interpretations of the EDE-Q (e.g., Parker et al., 2016; Liskowski et al., 2023) were not tested. Those included were based on models receiving the greatest empirical support (see Jenkins & Rienecke, 2022) and three full-item models showed similar fit, suggesting some redundancy of minor amendments. Similarly, several models did not converge, which could suggest that the data may not have contained enough information about the models (see Li, 2016), or that the models may have been underidentified due to the relatively few items per factor (but see Yang-Wallentin et al., 2010). Although practical and statistical differences are likely to be minimal (Riese et al., 1993), the models in the current study were identified

by fixing residual variances as opposed to fixing one factor loading (e.g., Grilo et al., 2013), and thus might offer different interpretations. Future studies might consider larger samples, particularly if looking at measurement invariance (e.g., Friborg et al., 2013; Rand-Giovannetti et al., 2020) or more complex (e.g., second-order) models.

The sample was recruited from a private treatment facility in the United States of America, in which patients must use insurance or pay out-of-pocket, thus possibly limiting the generalizability of our findings. In an effort to enhance generalizability, we included individuals who did not identify with the female gender (assigned at birth) although this could have introduced bias in the sample. However, whilst the EDE-Q may perform differently in samples of men or ethnic minorities, this appears to be less biased when using the EDE-Q7 (McEntee et al., 2021; see also Marek et al., 2023). Only those identified as underweight (BMI ≤19.0) were included, and therefore cases of AN and atypical AN at higher weights were not included; further studies might wish to consider a broader definition in psychometric work. Diagnoses were made according to clinical evaluations rather than standardized diagnostic interviews, which could have produced different functional psychiatric diagnoses (e.g., see Rettew et al., 2009). Finally, information on socioeconomic status was not available for this sample.

Conclusions

The current study evaluated several versions of the EDE-Q in a sample of women with AN, finding that a brief version (Grilo et al., 2013) showed excellent fit and three full-item versions (Friborg et al., 2013; Peterson et al., 2007; Phillips et al., 2018) demonstrated acceptable fit. The current findings suggest that the EDE-Q shows structural validity in this sample although interpretation based on the 'Original' factor structure seems unwarranted. Studies in this area might wish to consider using an appropriate full-item interpretation of the

EDE-Q (e.g., Friborg et al., 2013), or a measure such as the EDE-Q7 (Grilo et al., 2013) if a briefer assessment is sufficient.

References

Aardoom, J. J., Dingemans, A. E., Slof Op't Landt, M. C. T., & Van Furth, E. F. (2012). Norms and discriminative validity of the Eating Disorder Examination Questionnaire (EDE-Q). *Eating Behaviors, 13*(4), 305–309. https://doi.org/10.1016/j.eatbeh.2012.09.002

Allen, K. L., Byrne, S. M., Lampard, A., Watson, H., Fursland, A. (2011). Confirmatory factor analysis of the Eating Disorder Examination-Questionnaire (EDE-Q). *Eating Behaviors, 12*(2), 143–151. https://doi.org/10.1016/j.eatbeh.2011.01.005

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed). Arlington, VA: American Psychiatric Association.

Berg, K. C., Peterson, C. B., Frazier, P., & Crow, S. J. (2012). Psychometric evaluation of the Eating Disorder Examination and Eating Disorder Examination-Questionnaire: A systematic review of the literature. *International Journal of Eating Disorders, 45*, 428-438. https://doi.org/10.1002/eat.20931

Binford, R. B., Le Grange, D., & Jellar, C. C. (2005). Eating Disorders Examination versus Eating Disorders Examination-Questionnaire in adolescents with full and partial-syndrome bulimia nervosa and anorexia nervosa. *International Journal of Eating Disorders, 37*(1), 44–49. https://doi.org/10.1002/eat.20062

Bollen, K. A. (1989). Structural equations with latent variables. Wiley.

Brown (2015). Confirmatory Factor Analysis for Applied Research (2nd ed.). Guilford Press.

Calugi, S., Milanese, C., Sartirana, M., El Ghoch, M., Sartori, F., Geccherle, E., Coppini, A., Franchini, C., & Dalle Grave, R. (2017). The Eating Disorder Examination Questionnaire: reliability and validity of the Italian version. *Eating and Weight Disorders, 22*, 509-514. https://doi.org/10.1007/s40519-016-0276-6

Carey, M., Kupeli, N., Knight, R., Troop, N. A., Jenkinson, P. M., & Preston, C. (2019). Eating disorder examination questionnaire (EDE-Q): Norms and psychometric properties in U.K. females and males. *Psychological Assessment, 31*, 839-850. https://doi.org/10.1037/pas0000703

Cooper, P. J., & Fairburn, C. G. (1993). Confusion over the core psychopathology of bulimia nervosa. *International Journal of Eating Disorders, 13*(4), 385–389. https://doi.org/10.1002/1098-108X(199305)13:4%3C385::AID-EAT2260130406%3E3.0.CO;2-W

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods, 22*(3), 541–562. https://doi.org/10.1037/met0000083

Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders, 16*, 363–370. https://doi.org/10.1002/1098-108X(199412)16:4%3C363::AID-EAT2260160405%3E3.0.CO;2-%23

Fairburn, C. G., & Beglin, S. J. (2008). Eating Disorder Examination Questionnaire (EDE-Q 6.0). In C. G. Fairburn (Ed.), *Cognitive Behavior Therapy and Eating Disorders* (pp. 309-313). New York: Guilford Press.

Fairburn, C. G., Cooper, Z., & Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: a "transdiagnostic" theory and treatment. *Behaviour Research and Therapy, 41*(5), 509–528. https://doi.org/10.1016/S0005-7967(02)00088-8

Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science, 3*(4), 484–501. https://doi.org/10.1177/2515245920951747

Forrest, L. N., Jacobucci, R. C., Grilo, C. M. (2022). Empirically determined severity levels for binge-eating disorder outperform existing severity classification schemes. *Psychological Medicine 52*, 685–695. https://doi.org/10.1017/S0033291720002287

Friborg, O., Reas, D. L., Rosenvinge, J. H., & Rø, Ø. (2013). Core pathology of eating disorders as measured by the Eating Disorder Examination Questionnaire (EDE-Q): the predictive role of a nested general (*g*) and primary factors. *International Journal of Methods in Psychiatric Research, 22*(3), 195-203. https://doi.org/10.1002/mpr.1389

Goel, N. J., Burnette, C. B., Weinstock, M., & Mazzeo, S. E. (2022). Eating Disorder Examination-Questionnaire: Evaluating factor structures and establishing measurement invariance with Asian/Hawaiian/Pacific Islander, Black, and White American college men. *International Journal of Eating Disorders*. https://doi.org/10.1002/eat.23696

Grilo, C. M., Henderson, K. E., Bell, R. L., & Crosby, R. D. (2013). Eating disorder examination-questionnaire factor structure and construct validity in bariatric surgery candidates. *Obesity Surgery, 23*(5), 657–662. https://doi.org/10.1007/s11695-012-0840-8

Grilo, C. M., Reas, D. L., Hopwood, C. J., & Crosby, R. D. (2015). Factor structure and construct validity of the Eating Disorder Examination-Questionnaire in college students: Further support for a modified brief version. *International Journal of Eating Disorders, 48*(3), 284–289. https://doi.org/10.1002/eat.22358

Hrabosky, J. I., White, M. A., Masheb, R. M., Rothschild, B. S., Burke-Martindale, C. H., & Grilo, C. M. (2008). Psychometric evaluation of the eating disorder examination-questionnaire for bariatric surgery candidates. *Obesity, 16*, 763-769. https://doi.org/10.1038/oby.2008.3

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

Jenkins, P. E., & Davey, E. (2020). The brief (seven-item) eating disorder examination questionnaire: Evaluation of a non-nested version in men and women. *International Journal of Eating Disorders*, *53*, 1809–1817. https://doi.org/10.1002/eat.23360

Jenkins, P. E., & Rienecke, R. D. (2022). Structural validity of the Eating Disorder Examination – Questionnaire (EDE-Q): A systematic review. *International Journal of Eating Disorders*, *55*, 1012–1030. https://doi.org/10.1002/eat.23721

Kelley, K. (2017). MBESS (Version 4.0.0 and higher) [computer software and manual]. Accessible from http://cran.r-project.org.

Kliem, S., Mößle, T., Zenger, M., Strauß, B., Brähler, E., & Hilbert, A. (2016). The Eating Disorder Examination-Questionnaire 8: a brief measure of eating disorder psychopathology (EDE-Q8). *International Journal of Eating Disorders*, *49*(6), 613–616. https://doi.org/10.1002/eat.22487

Klimek, P., Convertino, A. D., Pennesi, J.-L., Gonzales, IV, M., Roesch, S. C., Nagata, J. M., & Blashill, A. J. (2021). Confirmatory factor and measurement invariance analyses of the Eating Disorder Examination Questionnaire in sexual minority men and women. *International Journal of Eating Disorders, 54*, 745-754. https://doi.org/10.1002/eat.23488

Korkmaz, S., Goksuluk, D., Zararsiz, G. (2014) MVN: An R package for assessing multivariate normality. *The R Journal*, *6*(2), 151–162. https://journal.r-project.org/archive/2014-2/korkmaz-goksuluk-zararsiz.pdf

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*, 606-613.

Laskowski, N. M., Halbeisen, G., Braks, K., Huber, T. J., & Paslakis, G. (2023). Exploratory graph analysis (EGA) of the dimensional structure of the eating disorder examination-questionnaire (EDE-Q) in women with eating disorders. *Journal of Psychiatric Research*. https://doi.org/10.1016/j.jpsychires.2023.05.063

Li, C.-H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods, 21*(3), 369–387. https://doi.org/10.1037/met0000093

Luce, K. H., Crowther, J. H., & Pole, M. (2008). Eating Disorder Examination Questionnaire (EDE-Q): Norms for undergraduate women. *International Journal of Eating Disorders, 41*(3), 273–276. https://doi.org/10.1002/eat.20504

Lydecker, J. A., White, M. A., & Grilo, C. M. (2017). Form and formulation: Examining the distinctiveness of body image constructs in treatment-seeking patients with binge-eating disorder. *Journal of Consulting and Clinical Psychology, 85*(11), 1095–1103. https://doi.org/10.1037/ccp0000258

McEntee, M.L., Serier, K.N., Smith, J.M., & Smith, J. E. (2021). The sum is greater than its parts: intersectionality and measurement validity of the Eating Disorder Examination Questionnaire (EDE-Q) in Latinx undergraduates in the United States. *Sex Roles, 84*, 102–111. https://doi.org/10.1007/s11199-020-01149-7

Machado, P. P. P., Grilo, C. M., Rodrigues, T. F., Vaz, A. R., & Crosby, R. D. (2020). Eating Disorder Examination – Questionnaire short forms: A comparison. *International Journal of Eating Disorders, 53*(6), 937–944. https://doi.org/10.1002/eat.23275

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods, 52*, 2287–2305. https://doi.org/10.3758/s13428-020-01398-0

Makowski, D., Ben-Shachar, M. S., Patil, I., & Lüdecke, D. (2019). Methods and Algorithms for Correlation Analysis in R. *Journal of Open Source Software, 5*(51), 2306. https://doi.org/10.21105/joss.02306

Marek, R. J., Ivezaj, V., Parikh, M. S., Jayade, M., Davila-Shiau, E., & Grilo, C. M. (2023). Factor structure and measurement invariance of the English- versus Spanish-language Eating Disorder Examination Questionnaire: Brief Form (S-EDE-Q-BF) in Hispanic/Latino/a/x persons seeking bariatric surgery. *Surgery for Obesity and Related Diseases, 19*(6), 576–584. https://doi.org/10.1016/j.soard.2022.12.015

Markon, K. E. (2019). Bifactor and hierarchical models: specification, inference, and interpretation. *Annual Review of Clinical Psychology, 15*, 51–69. https://doi.org/10.1146/annurev-clinpsy-050718-095522

Meczekalski, B., Podfigurna-Stopa, A., & Katulski, K. (2013). Long-term consequences of anorexia nervosa. *Maturitas, 75*(3), 215-220. https://doi.org/10.1016/j.maturitas.2013.04.014

Meule, A. (2021). Reconsidering the use of cut-off scores for the Eating Disorder Examination–Questionnaire. *Eating Disorders, 29*(5), 480–484. https://doi.org/10.1080/10640266.2019.1678981

Mond, J. M., & Hay, P. J. (2011). Dissatisfaction versus over-evaluation in a general population sample of women. *International Journal of Eating Disorders, 44*(8), 721–726. https://doi.org/10.1002/eat.20878

Moshagen, M., & Erdfelder, E. (2016). A new strategy for testing structural equation models. *Structural Equation Modeling, 23*, 54–60. https://doi.org/10.1080/10705511.2014.950896

Nagata, J. M., Compte, E. J., McGuire, F. H., Brown, T. A., Lavender, J. M., Murray, S. B., Capriotti, M. R., Flentje, A., Lubensky, M. E., Lunn, M. R., & Obedin-Maliver, J.

(2023). Investigating the factor structure and measurement invariance of the Eating Disorder Examination-Questionnaire (EDE-Q) in a community sample of gender minority adults from the United States. *International Journal of Eating Disorders*. https://doi.org/10.1002/eat.23978

O'Connor, S. M., Hazzard, V. M., & Zickgraf, H. F. (2022). Exploring differential item functioning on eating disorder measures by food security status. *Eating and Weight Disorders, 27,* 1449–1455. https://doi.org/10.1007/s40519-021-01289-z

Ortiz, S. N., Forrest, L. N., Kinkel-Ram, S. S., Jacobucci, R. C., Smith, A. R. (2021). Using shape and weight overvaluation to empirically differentiate severity of other specified feeding or eating disorder. *Journal of Affective Disorders, 295*, 446–452. https://doi.org/10.1016/j.jad.2021.08.049

Parker, K., Mitchell, S., O'Brien, P., & Brennan, L. (2016). Psychometric evaluation of disordered eating measures in bariatric surgery candidates. *Obesity Surgery, 26,* 563-575. https://doi.org/10.1007/s11695-015-1780-x

Passi, V. A., Bryson, S. W., & Lock, J. (2003). Assessment of eating disorders in adolescents with anorexia nervosa: self-report questionnaire versus interview. *International Journal of Eating Disorders, 33,* 45–54. https://doi.org/10.1002/eat.10113

Pennings, C., & Wojciechowski, F. L. (2004). Kort instrumenteel De Eating Disorder Examination Questionnaire (EDE-Q): Nederlandse normscores voor anorexiapatiënten en een nieteetstoornis controlegroep [The Eating Disorder Questionnaire (EDE-Q): Netherlands normative scores for anorexic patients and a non-eating disorder control group]. *Gedragstherapie, 37*(4), 293–301.

Peterson, C. B., Crosby, R. D., Wonderlich, S. A., Joiner, T., Crow, S. J., Mitchell, J. E., Bardone-Cone, A. M., Klein, M., & Le Grange, D. (2007). Psychometric properties of the eating disorder examination-questionnaire: Factor structure and internal

consistency. *International Journal of Eating Disorders, 40*, 386-389. https://doi.org/10.1002/eat.20373

Phillips, K. E., Jennings, K. M., & Gregas, M. (2018). Factor structure of the eating disorder examination-questionnaire in a clinical sample of adult women with anorexia nervosa. *Journal of Psychosocial Nursing and Mental Health Services, 56*, 33-39. https://doi.org/10.3928/02793695-20180108-03

Rand-Giovannetti, D., Cicero, D. C., Mond, J. M., & Latner, J. D. (2020). Psychometric properties of the Eating Disorder Examination-Questionnaire (EDE-Q): A confirmatory factor analysis and assessment of measurement invariance by sex. *Assessment, 27*, 164-177. https://doi.org/10.1177/1073191117738046.

Rettew, D. C., Lynch, A. D., Achenbach, T. M., Dumenci, L., & Ivanova, M. Y. (2009). Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *International Journal of Methods in Psychiatric Research, 18*(3), 169–184.

Revelle, W. (2021). *psych: Procedures for Personality and Psychological Research*. Northwestern University, Evanston. R package version 2.1.9.

Riese, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114*(3), 552–566. https://doi.org/10.1037/0033-2909.114.3.552

Rø, Ø., Reas, D. L., & Stedal, K. (2015). Eating Disorder Examination Questionnaire (EDE-Q) in Norwegian adults: discrimination between female controls and eating disorder patients. *European Eating Disorders Review, 23*(5), 408–412. https://doi.org/10.1002/erv.2372

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514. https://doi.org/10.1007/BF02296192

Schmitt, T. A., Sass, D. A., Chappelle, W., & Thompson, W. (2018). Selecting the "best" factor structure and moving measurement validation forward: An illustration. *Journal of Personality Assessment, 100*(4), 345–362. https://doi.org/10.1080/00223891.2018.1449116

Shi, D., DiStefano, C., McDaniel, H. L., & Jiang, Z. (2018). Examining Chi-Square Test Statistics under conditions of large model size and ordinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(6), 924–945. https://doi.org/10.1080/10705511.2018.1449653

Stice, E., Marti, C. N., & Rohde, P. (2013). Prevalence, incidence, impairment, and course of the proposed DSM-5 eating disorder diagnoses in an 8-year prospective community study of young women. *Journal of Abnormal Psychology*, *122*(2), 445-457. https://doi.org/10.1037/0030679

Surgenor, L. J., & Maguire, S. (2013). Assessment of anorexia nervosa: an overview of universal issues and contextual challenges. *Journal of Eating Disorders, 1*, 29. https://doi.org/10.1186/2050-2974-1-29

Widaman, K. F., & Revelle, W. (2023). Thinking thrice about sum scores, and then some more about measurement and analysis. *Behavior Research Methods*, *55*, 788-806. https://doi.org/10.3758/s13428-022-01849-w

Wolk, S. L., Loeb, K. L., & Walsh, B. T. (2005). Assessment of patients with anorexia nervosa: Interview versus self-report. *International Journal of Eating Disorders*, *37*, 92–99. https://doi.org/10.1002/eat.20076

Wothke, W. (1993). Nonpositive definite matrices in structural modelling. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp. 256–293). Newbury Park, CA: Sage.

Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling, 17*(3), 392–423. https://doi.org/10.1080/10705511.2010.489003

Table 1. Summary of included models.

Model (Authors)	Number of items / factors	Scales	Items (from Fairburn & Beglin, 2008)
Model 1 (Fairburn & Beglin, 2008)	22 / 4	Dietary Restraint Eating Concern Shape Concern Weight Concern	1–5 7,9,19,20,21 6,8,23,10,26,27,28,11 22,24,8,12,25
Model 2 (Friborg et al., 2013)	22 / 4	S1 (‘Shape and Weight Concern’) S2 (‘Preoccupation and Restriction’) S3 (‘Dietary Restraint’) S4 (‘Eating Shame’)	6,10–12, 22–28 2,5,7,8,9 1,3,4 19–21
Model 3 (Peterson et al., 2007)	22 / 3	Dietary Restraint Eating Concern Shape/Weight Concern	1-5 7,9,19,20,21 8,6,10,11,12,22–28
Model 4 (Grilo et al., 2013)	7 / 3	Restraint Body Dissatisfaction Shape/Weight Overvaluation	1,3,4 22,23 25,26
Model 5 (Phillips et al., 2018)	22 / 4	F1 F2 F3 F4	7–11,22–24 25–28 1–5,6,12 19–21
Model 6 (Pennings & Wojciechowski, 2004)	22 / 1	Global	1–12, 19–28
Model 7 (EDE-Q7-Bifactor)	7 / 4	Restraint Body Dissatisfaction Shape/Weight Overvaluation Global	1,3,4 22,23 25,26 All 7 items
Model 8 (Friborg-Bifactor)	22 / 5	S1 S2 S3 S4 Global	6,10–12, 22–28 2,5,7,8,9 1,3,4 19,20,21 All 22 items
Model 9 ($Friborg_{HO}$; Friborg et al., 2013)	22 / 4	S1 S2 S3 S4 Global	6,10–12, 22–28 2,5,7,8,9 1,3,4 19,20,21 All 22 items
Model 10 ($Phillips_{HO}$; not previously described)	22 / 5	F1 F2 F3 F4 Global	7–11,22–24 25–28 1–5,6,12 19–21 All 22 items

Table 2. Fit statistics and internal consistency reliability (Omega) for alternative factor structures of the EDE-Q in a sample of women with AN (N = 479).

Model	χ^2	df	CFI	TLI	SRMR	RMSEA (90% CIs)	Omega
Suggested cutoff	-	-	≥0.95	≥0.95	≤0.08	<0.10	-
Model 1 (‘Original’)	Did not converge		-	-	-	-	-
Model 2 (‘Friborg’)	2531.68	203	0.948	0.941	0.066	0.120 (0.115 – 0.124)	S1 = 0.95 S2 = 0.87 S3 = 0.90 S4 = 0.71
Model 3 (‘Peterson’)	3090.88	206	0.935	0.928	0.074	0.132 (0.128 – 0.136)	DR = 0.91 EC = 0.78 WSC = 0.96
Model 4 (‘EDE-Q-7’)	52.03	11	0.998	0.997	0.021	0.068 (0.050 – 0.087)	DR = 0.90 OWS = 0.95 BD = 0.88
Model 5 (‘Phillips’)	2363.30	203	0.952	0.945	0.064	0.115 (0.111 – 0.119)	F1 = 0.92 F2 = 0.91 F3 = 0.93 F4 = 0.71
Model 6 (‘One-factor’)	3580.73	209	0.924	0.917	0.086	0.142 (0.138 – 0.146)	0.96

Model 7: 'EDE-Q7-Bifactor'	Did not converge		-	-	-	-	-
Model 8: 'Friborg-Bifactor'	Did not converge		-	-	-	-	-
Model 9: $Friborg_{HO}$	2582.76	205	0.947	0.940	0.067	0.120 (0.116 – 0.124)	g = 0.90
Model 10: $Phillips_{HO}$	2375.65	205	0.951	0.945	0.065	0.115 (0.111 – 0.119)	g = 0.94

Note. Phillips F4 and Friborg S4 are comprised of the same items. Fit statistics for the Original model and bifactor models are not presented as the solutions did not converge (see text). CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; SRMR = Standardized Root Mean Square Residual; RMSEA = Root Mean Square Error of Approximation; CIs = confidence intervals; DR = Dietary Restraint; EC = Eating Concern; WSC = Weight and Shape Concern; OWS = Overvaluation of Weight and Shape; BD = Body Dissatisfaction; g = Global scale.

Table 3. Correlations (Spearman's ρ) between summed scales of the EDE-Q models (Friborg, Peterson, EDE-Q7, Phillips) and age, BMI, disordered eating behaviours, and depression

	Age	BMI	OBE	Self-induced vomiting	Laxative use	Excessive exercise	PHQ-9
Friborg (Model 2)							
S1	0.02	0.12*	0.09	0.26***	0.27***	0.35***	0.55***
S2	0.0004	0.10	0.22***	0.34***	0.29***	0.40***	0.59***
S3	0.02	0.05	0.10	0.26***	0.22***	0.49***	0.49***
S4	-0.04	0.08	0.21***	0.32***	0.23***	0.36***	0.58***
Peterson (Model 3)							
DR	0.02	0.08	0.12*	0.30***	0.27***	0.46***	0.54***
EC	-0.03	0.07	0.27***	0.34***	0.25***	0.38***	0.60***
WSC	0.02	0.12	0.10	0.27***	0.27***	0.37***	0.57***
EDE-Q7 (Model 4)							
DR	0.02	0.05	0.10	0.26***	0.22***	0.49***	0.49***
OWS	-0.0001	0.13*	0.11	0.21***	0.17***	0.26***	0.50***
BD	0.07	0.16**	0.12*	0.25***	0.20***	0.27***	0.46***
Phillips (Model 5)							
F1	0.0009	0.10	0.20***	0.31***	0.27***	0.39***	0.59***
F2	0.06	0.06	0.10	0.22***	0.22***	0.27***	0.48***
F3	0.02	0.10	0.11	0.30***	0.28***	0.46***	0.56***
F4	-0.04	0.08	0.21***	0.32***	0.23***	0.36***	0.58***

*p < .05, **p < .01, ***p < .001, Ns = 690–805

Note. Phillips F4 and Friborg S4 are comprised of the same items. BMI = body mass index; OBE = objective binge episodes; PHQ-9 = Patient Health Questionnaire-9; DR = Dietary Restraint; EC = Eating Concern; WSC = Weight and Shape Concern; OWS = Overvaluation of Weight and Shape; BD = Body Dissatisfaction.