


Psychometric Properties of the Eating Disorder Examination–Questionnaire (EDE-Q): A Confirmatory Factor Analysis and Assessment of Measurement Invariance by Sex

Assessment
1–14
© The Author(s) 2017
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1073191117738046
journals.sagepub.com/home/asm


Devin Rand-Giovanetti¹, David C. Cicero¹, Jonathan M. Mond^{2,3},
and Janet D. Latner¹

Abstract

The original, theoretically derived factor structure of the Eating Disorder Examination–Questionnaire (EDE-Q) has received limited empirical support and there is no consensus on an appropriate alternative. Moreover, there is a paucity of data on the factor structure of the EDE-Q across sexes. The goals of the current study were to evaluate models of the EDE-Q factor structure and to assess the best-fitting model for differences by sex. Twelve models were compared using confirmatory factor analysis in a sample of 940 undergraduates. Confirmatory factor analysis did not support the original factor structure. A four-factor model fit the data reasonably well with factors corresponding to themes of (a) dietary restraint, (b) preoccupation and restriction, (c) weight and shape concern, and (d) eating shame. The EDE-Q was found to be invariant by sex across all factors except Factor 3. The implications of these findings are discussed.

Keywords

eating disorders, confirmatory factor analysis, Eating Disorder Examination–Questionnaire, measurement invariance, sex, psychometrics

The Eating Disorder Examination (EDE; Fairburn & Cooper, 1993) structured clinical interview is one of the most widely used measures for the assessment of eating disorders (Guest, 2000). The Eating Disorder Examination–Questionnaire (EDE-Q; Fairburn, 2008b; Fairburn & Beglin, 1994) is a self-report version of the EDE commonly used for research and clinical purposes (Berg, Peterson, Frazier, & Crow, 2012; Mond, Hay, Rodgers, Owen, & Beumont, 2004a). The EDE-Q generates four rationally derived subscale scores (Restraint, Eating Concern, Weight Concern, and Shape Concern) and a global score (the mean of the subscale scores) designed to measure the occurrence and severity of eating disorder features. A substantial body of research has supported the internal consistency and temporal stability of the EDE-Q (Berg et al., 2012; Luce & Crowther, 1999; Mond et al., 2004a; Mond, Hay, Rodgers, Owen, & Beumont, 2004b).

The subscales of the EDE (and subsequently the EDE-Q) were developed based on themes pulled from unstructured interviews with eating disorder patients and a review of the literature (Cooper & Fairburn, 1987). In the original validation of the EDE, the authors stated that the subscales were

rationally derived and acknowledged that some features of the subscales were structured contrary to empirical evidence (Cooper, Cooper, & Fairburn, 1989). The authors noted, for example, that items in the Shape Concern and Weight Concern subscales were very highly correlated with one another in both the clinical and control samples, but chose to retain them as separate subscales because it was deemed premature to assume that they may be combined (Cooper et al., 1989). While this may have been true at the time of the original validation study, there is now considerable empirical evidence suggesting that these factors are not distinct (e.g., Hilbert, Tuschen-Caffier, Karwautz, Niederhofer, & Munsch, 2007; White, Haycraft, Goodwin, & Meyer, 2014). However,

¹University of Hawai'i at Mānoa, Honolulu, HI, USA

²University of Tasmania, Launceston, Tasmania, Australia

³Western Sydney University, Campbelltown, New South Wales Australia

Corresponding Author:

Devin Rand-Giovanetti, Department of Psychology, University of Hawai'i at Mānoa, 2530 Dole Street, Sakamaki C400, Honolulu, HI 96822, USA.
Email: drandgio@hawaii.edu

no subsequent versions of the EDE-Q have been changed to reflect this new evidence. Additionally, some items that did not correlate well with a given subscale were retained because they were deemed to be thematically alike. Although this approach may be appropriate for interpreting individual items descriptively, it can become problematic when the items are used to produce a mean subscale score that ostensibly reflects a single facet of a construct. Most empirical research has not supported the rationally derived subscale structure of the EDE and EDE-Q. Over 20 studies have investigated the factor structure of the EDE-Q during the past decade and all but 2 have failed to find support for the original structure (Table 1).

As seen in Table 1, several studies have employed both exploratory and confirmatory factor analytic approaches to assess the EDE-Q, in a range of populations. Results have yielded two-, three-, and four-factor models. Among studies that included all 22 original subscale items in their model, 7 employed an exploratory factor analysis (EFA) to examine the latent constructs underlying the measured variables. The resultant models produced both three-factor (Hilbert et al., 2007; Hilbert et al., 2012; Peterson et al., 2007; White et al., 2014) and four-factor solutions (Aardoom et al., 2012; Becker et al., 2010; Friborg et al., 2013). As aforementioned, several of these analyses found that most of the items from the Shape and Weight Concern subscales loaded together in the best-fitting solution (e.g., Hilbert et al., 2007; White et al., 2014). Prominent theories of body image (e.g., self-discrepancy theory, objectification theory) conceptualize body dissatisfaction as encompassing shape and weight concerns within a relatively unified construct (Cash, 2012; Vartanian, 2012). This suggests that the combination of these factors is a theoretically, as well as empirically, supported solution.

Several additional studies have used confirmatory factor analysis (CFA) to test the fit of the original and EFA- or theoretically-derived alternative models (Barnes et al., 2012; Giovazolias et al., 2013; Penelo et al., 2013). CFA has the advantage of testing a priori hypotheses concerning factor structure. As with the exploratory studies, these CFAs have yielded varied results that support a range of different factor structures (Table 1). Most of these CFAs, however, have each examined only a subset of the currently available models.

In an effort to improve the fit indicators of their model, several authors have removed items that did not load clearly onto any one factor (Allen et al., 2011; Calugi et al., 2016; Carrard et al., 2015; Chan & Leung, 2015; Darcy et al., 2013; Grilo et al., 2010; Grilo et al., 2013; Grilo et al., 2015; Hrabosky et al., 2008; Kliem et al., 2016; Parker et al., 2015, 2016; Wade et al., 2008). These reduced-item solutions yielded one-, two-, three-, and four-factor models that all supported the robustness of a model comprised primarily or exclusively of a single Weight/Shape Concerns subscale.

In some cases, the final, best-fitting model included as few as seven total items (e.g., Grilo et al., 2013). While this approach can greatly improve the fit indicators, it may also limit the scale's content validity, which could affect the relationship of the scale's scores with other variables. Additionally, the reduced-item CFA studies have utilized only a small portion of the available models. No consensus has yet been reached on the most appropriate factor structure for the EDE-Q. There is a need for a comprehensive analysis of all available full-measure and reduced-item models in the same sample.

Further research is also needed to elucidate sex differences in the EDE-Q factor structure, particularly given recent evidence suggesting increases in the population prevalence of eating pathology in males (Hudson, Hiripi, Pope, & Kessler, 2007; Mitchison, Hay, Slewa-Younan, & Mond, 2014; Mitchison & Mond, 2015). Evidence suggests that there are key differences in the presentation of disordered eating behavior among men when compared with women (Mitchison & Mond, 2015; Stanford & Lemberg, 2012). In particular, muscularity-oriented eating pathology is more prominent among men (Mitchison & Mond, 2015). This may pose a problem for the interpretation of research employing measures that were developed with female populations in mind, including the EDE and EDE-Q (Mitchison & Mond, 2015).

To date, only two studies have tested the factor structure of the EDE-Q separately in men and women. In a study of college athletes, Darcy et al. (2013) performed EFAs for four groups of college students: male and female athletes and non-athletes. The factor structures produced by these analyses suggested that the factor loadings for male nonathletes were divergent from the other groups examined. In a CFA conducted by Chan and Leung (2015), support was found for the fit of a one-factor, eight-item model in women but not men. Three further studies have employed measurement invariance analysis, an assessment of model uniformity across groups (Chen, 2008), to examine whether men and women interpret and respond to the questions in same manner. In their research on Mexican children and adolescents, Penelo et al. (2013) found support for measurement invariance across sexes using a Spanish translation of the EDE-Q. Grilo et al. (2015) and Kliem et al. (2016) also found evidence of measurement invariance across sexes for their reduced-item models in samples of American and German adults, respectively. To our knowledge, no research has assessed EDE-Q measurement invariance across sexes using the full measure in a sample of English-speaking adults.

The primary goal of the current study was to evaluate and compare different models of the EDE-Q factor structure in a single sample. Models were selected that (a) had a unique, complete published factor structure and (b) included all 22 subscale items in the final model. Twelve models were identified that met these criteria:

Table 1. Review of EFA and CFA Studies of the EDE-Q.

Authors (year)	EDE-Q version	Participants	Methods	Results
<i>Full-item models</i>				
Hilbert et al. (2007)	German translation based on EDE-Q 4.0 and 5.0	<ul style="list-style-type: none"> ED (n = 214) Subthreshold ED (n = 32) Psychiatric comparison (n = 51) Nonclinical comparison (n = 409) 89.1% female 	EFA (principal component analysis)	Did not replicate original model; supported three-factor model with all subscale items
Peterson et al. (2007)	Original EDE-Q, 36 items	<ul style="list-style-type: none"> Threshold and subthreshold BN (N = 203; 100% female) 	EFA (principal axis analysis)	Did not replicate original model; produced three- and four-factor model, most support for three-factor with all subscale items
Becker et al. (2010)	EDE-Q 5.2, 28 items, Fijian adaptation	<ul style="list-style-type: none"> Fijian adolescents aged 15-20 years (N = 523; 100% female) 	EFA (principal axis factoring), requested four-factor solution	Did not replicate original model; supported four-factor model with all subscale items <i>Not included in current analyses, item loadings not published</i>
Villarroel, Penelo, Portell, and Raich (2011)	EDE-Q 4.0, 38 items, Spanish adaptation	<ul style="list-style-type: none"> College students (N = 708; 100% female) 	CFA of original four-factor model	Supported original four-factor model
Aardoom, Dingemans, Slop Op't Landt, and Van Furth (2012)	Original EDE-Q, 36 items, Dutch translation	<ul style="list-style-type: none"> ED (AN, BN, BED, EDNOS) treatment-seeking (N = 935; 100% female) 	EFA (principal component analysis)	Did not replicate original model; supported four-factor model with all subscale items
Barnes, Prescott, and Muncer (2012)	EDE-Q 6.0, 28 items	<ul style="list-style-type: none"> ED treatment-seeking (n = 166; 95.8% female) College students (n = 403; 91.8% female) 	CFA of original four-factor, one-factor all items, Peterson et al.'s three-factor model; groups analyzed together	Supported Peterson et al.'s three-factor model; model fit both groups
Franko et al. (2012)	Original EDE-Q, 36 items	<ul style="list-style-type: none"> Latina college students (N = 173; 100% female) 	CFA of original four-factor	Supported original four-factor model
Hilbert, de Zwaan, and Braehler (2012)	EDE-Q 6.0, 28 items, German translation	<ul style="list-style-type: none"> Community (N = 2,520, 53.7% female) 	EFA (principal component analysis)	Supported three-factor model with all subscale items
Friborg, Reas, Rosenvinge, and Rø (2013)	EDE-Q 6.0, 28 items, Norwegian translation	<ul style="list-style-type: none"> Community (N = 1,076; 100% female) 	EFA on half of sample, CFA on other half of sample for one-factor all items, original four-factor, EFA three-factor, EFA four-factor, nested four-factor w/general factor	Did not replicate original model; supported nested four-factor model with general factor using all subscale items
Giozoulis, Tsaousis, and Vallianatou (2013)	Original EDE-Q, 36 items, Greek translation	<ul style="list-style-type: none"> College students (N = 500; 100% female) 	CFA of one-factor with all items, original four-factor, Hilbert et al.'s three-factor model, Peterson et al.'s three-factor model	Supported Peterson et al.'s three-factor model
Penelo, Negrete, Portell, and Raich (2013)	EDE-Q 4.0, 38 items, Spanish translation	<ul style="list-style-type: none"> Children, aged 11-18 years (N = 2,928; 52.7% female) 	CFA of original four-factor, three-factor retaining Restraint and Eating, two-factor retaining Restraint, one-factor all items	Supported two-factor model with all subscale items <i>Not included in current analyses as model is identical to included Allen's two-factor</i>
White et al. (2014)	EDE-Q 6.0, 28 items	<ul style="list-style-type: none"> Adolescents aged 14-18 years (N = 917; 56.9% female) 	CFA on half of sample for original four-factor; EFA on half of sample	Did not replicate original model; supported three-factor model with all subscale items
<i>Reduced-item models</i>				
Hrabosky et al. (2008)	Original EDE-Q, 36 items	<ul style="list-style-type: none"> Obese bariatric surgery candidates (N = 337; 83.3% female) 	EFA on half of sample; CFA on other half of sample for EFA-derived four-factor model (analysis included both subscale and nonsubscale items)	Did not replicate original model; supported four-factor model with 12-items (three subscale items factors, one nonsubscale items factor) <i>Included Grilo, Henderson, Bell, and Crosby's (2013) modified three-factor version of model in current analyses which excludes nonsubscale factor</i>
Allen, Byrne, Lampard, Watson, and Fursland (2011)	Not specified	<ul style="list-style-type: none"> ED (AN, BN, EDNOS) treatment-seeking (n = 228; 100% female) Community (n = 227; 100% female) 	CFA of original four-factor, three-factor, two-factor, one-factor w/all items, Wade et al.'s brief one-factor; groups analyzed separately	Supported Wade et al.'s brief one-factor model with eight items; model fit both groups
Darcy, Hardy, Crosby, Lock, and Peebles (2013)	EDE-Q 6.0, 28 items	<ul style="list-style-type: none"> Male college students (n = 229) Female college students (n = 429) Male college athletes (n = 432) Female college athletes (n = 544) 	CFA of original four-factor; groups analyzed separately; EFA (promax oblique rotation); groups analyzed separately	Did not replicate original model; supported three-factor models for all groups except male nonathletes (18-21 items), two-factor model supported for male nonathletes (19 items)

(continued)

Table I. (continued)

Authors (year)	EDE-Q version	Participants	Methods	Results
Grilo et al. (2013)	Original EDE-Q, 36 items	• Obese bariatric surgery candidates ($N = 174$; 75.2% female)	CFA of original four-factor, Hrabosky et al.'s brief three-factor model, ^b Grilo et al.'s (2010) ^c brief three-factor model	Supported Grilo et al.'s (2010) ^c brief three-factor model with seven items
Carrard, My Lien Rebetz, Mobbs, and Van der Linden (2015)	EDE-Q 6.0, 28 items, French translation	• BED treatment-seeking ($n = 116$; 100% female) • Community ($n = 161$; 100% female)	CFA of Peterson et al.'s three-factor model, Grilo et al.'s (2010) ^c brief three-factor model, Wade et al.'s ^a brief one-factor model; groups analyzed separately	Supported Grilo et al.'s (2010) ^c three-factor model with seven items; model fit both groups
Chan and Leung (2015)	EDE-Q 6.0, 28 items	• College students ($N = 310$, 54.2% female)	CFA of Peterson et al.'s three-factor, two-factor model, one-factor w/all items, Wade et al.'s ^a brief one-factor model	Supported Wade et al.'s ^a brief one-factor model with eight items
Grilo, Reas, Hopwood, and Crosby (2015)	Original EDE-Q, 36 items	• College students ($N = 801$; 71.4% female)	CFA of original four-factor, Hrabosky et al.'s brief three-factor model, ^b Grilo et al.'s (2010) ^c brief three-factor model	Supported Grilo et al.'s (2010) ^c brief three-factor model with seven items
Kliem et al. (2016)	EDE-Q 6.0, 28 items, German translation	• Community ($N = 2,508$, 53.2% female)	CFA of a one-factor EDE-Q developed by authors ^d comparing a one-factor and four-factor eight-item model with a higher order factor	Supported four-factor model with eight items
Parker, Mitchell, O'Brien, and Brennan (2015)	EDE-Q 6.0, 28 items	• Post bariatric surgery patients ($N = 108$, 80.6% female)	CFA of the original factor structure; EFA	CFA did not support original model; EFA supported a four-factor with 14 item model
Calugi et al. (2016)	EDE-Q 6.0, 28 items, Italian translation	• ED treatment-seeking ($N = 264$, 97.3% female)	CFA of one-factor with all items, original four-factor, Grilo et al.'s (2010) ^c brief three-factor model	Supported Grilo et al.'s (2010) ^c brief three-factor model with seven items
Gideon et al. (2016)	EDE-Q 6.0, 28 items	• ED treatment-seeking ($N = 489$, 90.2% female)	EFA (Principal component analysis)	Did not replicate original model; supported five-factor model with 12 item model <i>Not included in current analyses as authors changed item content</i>
Parker, Mitchell, O'Brien, and Brennan (2016)	EDE-Q 6.0, 28 items	• Bariatric surgery candidates ($N = 405$, 79.3% female)	CFA of Allen et al.'s three-factor model; EFA	CFA did not support Allen et al.'s three-factor model, EFA supported a four-factor model with 14 item model

Note. AN = anorexia nervosa; BED = binge eating disorder; BN = bulimia nervosa; CFA = confirmatory factor analysis; ED = eating disorder; EDE = Eating Disorder Examination; EDNOS = eating disorder not otherwise specified; EFA = exploratory factor analysis; EDE-Q = Eating Disorder Examination–Questionnaire.

^aWade, Byrne, and Bryant-Waugh (2008) conducted an EFA on the EDE in a sample of female twin children and produced a brief eight-item, one-factor model comprised of items from the Shape/Weight Concern subscales. Not included in this table because they did not assess the EDE-Q. ^bHrabosky et al.'s model used in these analyses includes the three factors from their four-factor model that contain subscale items, but excludes the factor with nonsubscale items. ^cGrilo et al. (2010) conducted an EFA and CFA on the EDE in a sample of BED patients and produced a brief seven-item, three-factor model. Not included in this table because they did not assess the EDE-Q. ^dItems in the shortened EDE-Q were selected based on achieving (a) optimal internal consistency, (b) unidimensionality, (c) fewer items than the original scale, and (d) an equal number of items from each of the four original subscales.

1. A one-factor model including all 22 items
2. Allen et al.'s (2011) two-factor model
3. Allen et al.'s three-factor model
4. Hilbert et al.'s (2007) three-factor model
5. Hilbert et al.'s (2012) three-factor model
6. Peterson et al.'s (2007) three-factor model
7. White et al.'s (2014) three-factor model
8. Aardoom et al.'s (2012) four-factor model
9. Fairburn et al.'s original four-factor model
10. Friborg et al.'s (2013) four-factor model
11. Peterson et al.'s four-factor model
12. Higher order model

Models that excluded subscale items or that included nonsubscale items were not included for comparative purposes because these models have different manifest variables (i.e., variables that are directly measured or observed) and different *numbers* of manifest variables, precluding

meaningful comparison of model fit. Additional analyses were, however, conducted to assess fit indices for the reduced-item models, as several of these models have demonstrated good fit in previous research. Nine such models were identified:

1. Wade et al.'s (2008) one-factor model
2. Darcy et al.'s (2013) two-factor model (male nonathlete)
3. Darcy et al.'s three-factor model (female nonathlete)
4. Darcy et al.'s three-factor model (male athlete)
5. Darcy et al.'s three-factor model (female athlete)
6. Grilo et al.'s adaptation of Hrabosky et al.'s three-factor model (Grilo et al., 2013; Grilo et al., 2015)
7. Grilo et al.'s (2010; Grilo et al., 2013; Grilo et al., 2015) three-factor model
8. Parker et al.'s (2015, 2016) four-factor model
9. Kliem et al.'s (2016) four-factor model (with a higher order general factor)

Based on findings from previous research, it was hypothesized, first, that support would not be found for Cooper and Fairburn's (1987) original model; and second, that the best-fitting model would include a factor containing items from both the Weight Concern and Shape Concern subscales. No hypotheses were made regarding the number of factors in the final model, given the heterogeneity of findings from previous research in this regard.

An additional aim of the current research was to evaluate measurement invariance by sex to determine whether the EDE-Q has the same psychometric properties in men and women. The paucity of existing evidence precluded any a priori hypotheses in this regard.

Method

Participants

Participants were 981 undergraduate psychology students recruited from a large Pacific public university who participated in exchange for partial completion of a course requirement or extra credit. Data were collected online during the fall 2013 and spring 2014 semesters. Response rates by semester cannot be calculated due to a limitation of the online survey system. On average, however, 95.8% of individuals who signed up for the survey system since 2013 have completed the questionnaire battery. Forty-one individuals had at least one missing data point and were excluded listwise from the analysis. The data were missing completely at random, Little's missing completely at random test, $\chi^2(295) = 268.329, p = .865$. This resulted in 940 participants being included in the analyses. Participants were 69.9% female, 29.6% male, and 0.1% other. They ranged in age from 16 to 48 years, $M(SD) = 20.34(3.74)$, and their mean body mass index (kg/m^2), calculated using self-reported height and weight, was $23.28(SD = 4.56)$. Participants self-identified as 50.7% Asian or Pacific Islander, 20.6% Caucasian, 16.3% biracial or multiracial, 5.3% Native Hawai'ian, Native American, or American Indian, 3.8% Hispanic, 1.0% Black, and 1.6% other. This research was approved by the University of Hawai'i Institutional Review Board and participants provided informed consent.

Measures and Procedures

The EDE-Q (6.0; Fairburn, 2008b) is a self-report measure consisting of 28 items assessing core features of eating disorder symptomatology. The EDE-Q produces two types of data: severity and frequency. Twenty-two items assessing severity are rated on a 7 point (0-6) forced-choice scale, with higher scores indicating greater severity. In the original model, these severity items are used to generate a global score and four subscale scores: Restraint (five items), Eating Concern (five items), Shape Concern (eight items),

and Weight Concern (five items). Six additional items assess the frequency of key behavioral features of eating disorders such as binge eating, self-induced vomiting, laxative misuse, diuretic misuse, and excessive exercise. The behavioral frequency items do not contribute to the subscale scores, but provide clinically useful information that may inform diagnostic and treatment decisions. All models tested in these analyses included only the subscale items.

Basic demographic information, including age, sex, self-reported height, and self-reported weight, was also collected.

Statistical Analyses

Confirmatory Factor Analyses Comparison of Model Fit. The first goal of the current research was to test the factor structure for the 12 EDE-Q models that include all subscale items. All model fitting was done with *Mplus* 7.3 software (Muthén & Muthén, 1998-2012). Models were specified with unweighted least squares mean and variance adjusted estimation and the "categorical" option in *Mplus*.¹ This method is appropriate for Likert-type scale data, which are ordered categorical variables (Muthén & Muthén, 1998-2012). Following Hu and Bentler's (1998) guidelines, three fit statistics were used to determine whether the models fit the data well: root mean square error of approximation (RMSEA) < 0.08, comparative fit index (CFI) > 0.95, and Tucker-Lewis index (TLI) > 0.95. The best fitting four-factor model was compared with the best fitting three-, two-, and one-factor models with a Satorra-Bentler χ^2 (SB χ^2) difference test using the "diffest" command (Satorra & Bentler, 2001). The best-fitting model was tested in a higher order model in which all the first-order factors loaded on a single higher order factor. This model was also used to test the measurement invariance of the scale between sexes.

Additionally, fit statistics were calculated for nine reduced item models. Because these models all include different items, their fit cannot be compared with each other or to the models that include all subscale items. The brief models were fit with the same specifications as the full models and fit statistics are provided for comparison with Hu and Bentler's (1998) guidelines.

Measurement Invariance Between Sexes. The second goal of the current research was to examine the measurement invariance of the EDE-Q in men and women. While the majority of our sample was female (70%), our sample of male participants ($n = 283$) was larger than the 200 person per group minimum suggested in Monte Carlo studies for the analysis of measurement invariance (Cheung & Rensvold, 2002). Thus, we had sufficient statistical power to evaluate measurement invariance. In all models, the individual items of the EDE-Q were the manifest variables. We used maximum likelihood estimation for the measurement invariance analyses for several reasons. First, Monte Carlo simulation studies have shown that least squares approaches are appropriate for

Table 2. Fit Statistics for Competing All-Item Models in the Full Sample.

Model	χ^2	df	RMSEA	90% CI	TLI	CFI
One-factor	2094.935	209	0.098	[0.094, 0.102]	0.909	0.918
Allen's two-factor	1728.734	208	0.088	[0.088, 0.092]	0.926	0.934
Allen's three-factor	1666.821	206	0.087	[0.083, 0.091]	0.929	0.936
Hilbert's (2007) three-factor	1703.274	206	0.088	[0.084, 0.092]	0.926	0.934
Hilbert's (2012) three-factor	1488.795	206	0.081	[0.078, 0.085]	0.937	0.944
Peterson's three-factor	1563.471	206	0.084	[0.080, 0.088]	0.934	0.941
White's three-factor	1696.221	206	0.088	[0.084, 0.092]	0.927	0.935
Aardoom's four-factor	1503.797	203	0.083	[0.079, 0.086]	0.936	0.943
Fairburn's four-factor	1653.581	203	0.087	[0.083, 0.091]	0.928	0.937
Friborg's four-factor	1339.707	203	0.077	[0.073, 0.081]	0.944	0.951
Peterson's four-factor	1587.684	203	0.085	[0.081, 0.089]	0.931	0.939
Higher order	1329.255	205	0.076	[0.072, 0.080]	0.951	0.945

Note. χ^2 = chi-square; df = degrees of freedom; RMSEA = root mean square error of approximation; CI = confidence interval; TLI = Tucker–Lewis index; CFI = comparative fit index.

Likert-type data with 2 to 5 response options, but that least squares and maximum likelihood estimation perform equally well with 6 to 7 response options (Rhemtulla, Brosseau-Liard, & Savalei, 2012). Second, in *Mplus*, measurement invariance analyses with least squares approaches for categorical variables require that all the response options are endorsed in all groups (Muthén & Muthén, 1998–2012). In the current research, the most extreme option was not chosen by men for some variables. One solution is to collapse across response options so that each response option is endorsed by both groups for each variable (e.g., recoding “7” responses into “6”). However, this approach does not seem appropriate for measurement invariance analyses because the finding that one group systematically avoided certain response options could itself be indicative of a lack of scalar or metric invariance. Third, the use of maximum likelihood estimation allows for the use of alternative tests for measurement invariance (i.e., McDonald's noncentrality index [NCI] and change in CFI) that were developed and tested within maximum likelihood estimations (McDonald, 1989; Meade, Johnson, & Braddy, 2008).

The data were fitted to three different measurement invariance models: configural, metric, and scalar. Configural invariance indicates whether there is the same pattern of factor loadings between groups. Metric invariance indicates whether the factor loadings are equivalent between groups. If a scale fails to display metric invariance, then it is likely that the scale is measuring a different construct in each group. Scalar invariance refers to the equivalence of the intercepts between groups. If the scale does not display scalar invariance, then the same score may represent a different level of pathology in one group than it does in another group.

In the configural invariance model, all the factor loadings and intercepts were allowed to load freely and differ between sexes. In the metric invariance model, the factor loadings were constrained to be equal between sexes, but

the intercepts were allowed to differ. In the scalar invariance model, the factor loadings and intercepts were constrained to be equal between sexes. The fit of the metric and scalar invariance models were compared with the fit of the configural model (Chen, 2008). If the scale had failed to display metric or scalar invariance, we planned to examine the modification indices to determine which loadings or intercepts were responsible for the lack of invariance (Byrne, Shavelson, & Muthén, 1989; Marsh & Hocevar, 1985; Sörbom, 1989; van de Schoot et al., 2013). Follow-up analyses were planned to focus on parameters with modification indices greater than 10.00, as recommended (Heene, Hilbert, Freudenthaler, & Bühner, 2012).

We used several tests to determine whether the scale was invariant. Following typical protocol for CFA, we reported the SB χ^2 difference test. Research has shown, however, that chi-square based likelihood ratio tests are problematic when comparing model fit, especially for tests of measurement invariance (Cheung & Rensvold, 2002). Thus, the SB χ^2 was supplemented with McDonald's (1989) NCI and change in CFI as suggested by Meade et al. (2008). Following the recommendations of Cheung and Rensvold (2002), the cutoffs of 0.020 for NCI and 0.010 for change in CFI were used. Mean scores for all items and factors of the best-fitting model were calculated separately for men and women to examine sex differences in responses.

Results

Full Measure Model Fit Comparisons

Friborg et al.'s (2013) four-factor model fit the data well according to the RMSEA (0.077) and CFI (0.951) statistics and nearly met criteria for TLI (0.944). None of the other models met criteria for good fit according to any of the fit statistics reported in Table 2. Friborg et al.'s (2013) four-factor

Table 3. Fit Statistics for Reduced-Item Models in the Full Sample.

Model	χ^2	df	RMSEA	90% CI	TLI	CFI
Wade's one-factor	271.859	20	0.115	[0.103, 0.128]	0.979	0.985
Darcy's two-factor (male nonathlete)	1094.738	134	0.087	[0.082, 0.092]	0.936	0.944
Darcy's three-factor (female nonathlete)	1099.733	149	0.082	[0.078, 0.087]	0.947	0.954
Darcy's three-factor (male athlete)	1168.038	186	0.075	[0.071, 0.079]	0.949	0.955
Darcy's three-factor (female athlete)	1286.388	167	0.084	[0.080, 0.089]	0.942	0.949
Grilo's three-factor	72.418	11	0.077	[0.060, 0.094]	0.976	0.987
Hrabosky's three-factor	118.322	24	0.064	[0.053, 0.076]	0.984	0.989
Parker's four-factor	290.058	71	0.057	[0.050, 0.064]	0.977	0.982
Kleim's four-factor (w/higher order)	116.097	16	0.099	[0.086, 0.113]	0.964	0.979

Note. χ^2 = chi-square; RMSEA = root mean square error of approximation; TLI = Tucker–Lewis index; CFI = comparative fit index; df = degrees of freedom; CI = confidence interval.

model fit significantly better than the best fitting three-factor model (Hilbert et al., 2012), $\chi^2(3) = 100.25, p < .001$; two-factor model (Allen et al., 2011), $\chi^2(5) = 236.340, p < .001$; and the one-factor model, $\chi^2(6) = 403.478, p < .001$. Taken together, these results suggest that Friborg et al.'s (2013) model is the best-fitting model in the total sample. If we examined the factor structure separately for women and men, the results were nearly identical to the full sample. In both samples, Friborg et al.'s (2013) four-factor model fit the data best according to all the test statistics, and fit significantly better than all other models as in the full sample.

We also tested whether a higher order model, in which the first-order factors loaded on a single higher order model, would fit just as well or significantly worse than the first-order model. For Friborg et al.'s (2013) first-order model, the four factors are allowed to correlate freely with each other (six total parameters). In the higher order model, the correlations among the factors are represented by the first-order factors loading on a higher order model (four total parameters). Thus, the higher order model is more restrictive and cannot fit better than the first-order model (McGartland Rubio, Berg-Weger, & Tebb, 2001). Since the higher order model is more parsimonious (i.e., it has two fewer parameters), we tested whether it fit significantly worse than the first-order model. The Friborg higher order model fit significantly worse than the Friborg first-order model, SB χ^2 difference test (2) = 25.652, $p < .001$.

Next, an exploratory post hoc analysis was conducted to assess the validity of Friborg et al.'s model. This was accomplished by examining the relationship between the factors and the EDE-Q's behavioral frequency items that do not contribute to the subscale scores. Following Mond, Hay, Rodgers, and Owen (2006), we created three dichotomous variables for eating pathology including, (a) binge eating once a week or more (i.e., >3 times in the 28 days), (b) purging with vomiting or laxative use once a week or more (i.e., >3 times in the past 28 days), and (c) excessive exercise 5 times a week or more (i.e., >19 times in the past 28 days). The bivariate relations among these three variables and the four factors were

all positive and statistically significant (all p 's < .001). In simultaneous regression analyses with binge eating, purging, and excessive exercise regressed on the four factors, Factor 4 was positively related to binge eating ($\beta = .36, p = .022$), purging ($\beta = .75, p = .009$), and excessive exercise ($\beta = .48, p = .011$). In contrast, Factor 3 was negatively associated with purging ($\beta = -.46, p = .023$) and excessive exercise ($\beta = -.33, p = .025$). Factor 1 was positively associated with excessive exercise ($\beta = .60, p < .001$). Factor 2 was not associated with any of the eating pathology behaviors.

Reduced-Item Model Analyses

As can be seen in Table 3, an analysis of the fit statistics for the reduced-item models revealed that several models fit the data well according to Hu and Bentler's (1998) guidelines. Darcy et al.'s (2013) three-factor model developed in female nonathletes and their three-factor model developed in male athletes both demonstrated good fit, as did Hrabosky et al.'s (2008) three-factor model, Grilo et al.'s (2013) three-factor model, and Parker et al.'s (2015, 2016) four-factor model. However, the fit of these brief models cannot be compared with the fit of the full models because they include different data and are nonnested.

Measurement Invariance Between Sexes

Measurement invariance analyses were conducted using Friborg et al.'s (2013) four-factor model. The configural model, in which the factor loadings and intercepts were free to vary between sexes, fit the data reasonably well (Table 4). The factor loadings were similar in both men and women (Table 5). The metric invariance model, in which the factor loadings were constrained to be equal in men and women, also fit the data reasonably well (Table 4). This model did not fit significantly worse than the configural model according to the NCI and change in CFI statistics, suggesting that the EDE-Q has metric invariance between sexes for Friborg et al.'s (2013) model. However, the scalar invariance model

Table 4. Model Fit for the Configural, Metric, and Scalar Invariance Model for Friborg et al.'s (2013) Four-Factor Model.

Model	χ^2	df	RMSEA	90% CI	TLI	CFI	SRMR	χ^2_{diff} (df)	p	Δ CFI	NCI
Configural ^a	1826.085	406	0.086	[0.082, 0.091]	0.827	0.848	0.069	—	—	—	—
Metric ^b	1870.324	424	0.085	[0.081, 0.089]	0.831	0.845	0.084	50.919 (18)	< .001	0.003	0.006
Scalar ^c	1963.182	442	0.086	[0.082, 0.090]	0.829	0.837	0.086	136.719 (36)	< .001	0.011	0.024
Scalar-modified ^d	1930.264	440	0.085	[0.081, 0.089]	0.832	0.840	0.085	103.978 (34)	< .001	0.008	0.016

Note. χ^2 = chi-square; RMSEA = root mean square error of approximation; TLI = Tucker–Lewis index; CFI = comparative fit index; SRMR = standardized root mean square residual; Δ CFI = change in CFI; NCI = McDonald's noncentrality index; df = degrees of freedom; CI = confidence interval.

^aConfigural model: No constraints on loadings or intercepts between groups. ^bMetric model: Loadings are constrained to be equal between groups.

^cScalar model: Loadings and intercepts are constrained to be equal between groups. ^dScalar-modified model: All loadings are constrained to be equal between groups. All intercepts except Items 6 and 10 are constrained to be equal between groups.

Table 5. Factor Loadings of the Configural Model, Mean (SD), and Mean Comparisons Between Men and Women.

Item	Factor loadings		M (SD)		Test statistic	
	Men	Women	Men	Women	t (971)	p
Factor 1			1.72 (2.09)	2.27 (2.05)	3.85	<.001
Restraint over eating	.81	.78	1.65 (2.37)	2.51 (2.42)	5.10	<.001
Food avoidance	.78	.83	1.93 (2.46)	2.47 (2.36)	3.22	<.001
Dietary rules	.84	.82	1.57 (2.34)	1.84 (2.25)	1.70	.089
Factor 2			0.47 (0.95)	0.87 (1.28)	10.25	<.001
Avoidance of eating	.51	.63	0.31 (1.01)	0.58 (1.31)	3.12	.002
Empty stomach	.56	.69	0.51 (1.46)	0.76 (1.55)	2.31	.020
Preoccupation with food, eating, calories	.62	.64	0.59 (1.52)	1.03 (1.89)	3.55	<.001
Preoccupation with shape and weight	.63	.78	0.49 (1.42)	0.91 (1.76)	3.69	<.001
Fear of losing control over eating	.68	.69	0.49 (1.40)	1.08 (1.96)	4.66	<.001
Factor 3			1.32 (1.36)	2.51 (1.74)	NA	NA
Flat stomach	.50	.53	2.14 (2.67)	3.88 (2.52)	NA	NA
Fear of weight gain	.57	.72	0.91 (1.88)	2.60 (2.61)	NA	NA
Feelings of fatness	.75	.78	1.68 (2.29)	3.17 (2.37)	9.06	<.001
Desire to lose weight	.68	.76	1.98 (2.60)	3.35 (2.55)	7.68	<.001
Importance of weight	.78	.83	1.00 (1.51)	1.79 (1.97)	6.09	<.001
Importance of shape	.74	.80	1.21 (1.60)	1.93 (1.97)	5.42	<.001
Reaction to prescribed weighing	.36	.59	0.33 (0.98)	1.19 (1.73)	7.88	<.001
Dissatisfaction with weight	.82	.90	1.23 (1.64)	2.46 (2.12)	8.71	<.001
Dissatisfaction with shape	.87	.91	1.44 (1.62)	2.57 (2.02)	8.34	<.001
Discomfort seeing body	.82	.89	1.03 (1.54)	2.19 (2.00)	8.78	<.001
Discomfort exposing body	.84	.83	1.36 (1.71)	2.43 (2.11)	7.53	<.001
Factor 4			0.29 (0.60)	0.83 (1.11)	73.66	<.001
Eating in secret	.41	.55	0.09 (0.38)	0.31 (0.79)	4.59	<.001
Guilt after eating	.80	.86	0.54 (1.20)	1.38 (1.81)	7.18	<.001
Social eating	.41	.71	0.27 (0.85)	0.81 (1.39)	6.09	<.001

fit significantly worse than the configural model according to all three model comparisons. The modification indices suggested that freeing the intercepts of Item 6, “Have you had a definite desire to have a *totally flat* stomach?” and Item 10, “Have you had a definite fear that you might gain weight?” would improve model fit. As shown in Table 4, a modified scalar invariance model in which all loadings and intercepts except the intercepts for Questions 6 and 10 were constrained to be equal between sexes fit just as well as the

configural invariance model according to the NCI and change in CFI statistics. This suggests that differences in intercepts between the sexes on Question 6 and Question 10 are responsible for the observed lack of scalar invariance.

Mean Comparisons Between Men and Women

Means were compared for each individual item and for factor scores. It was expected that women would have higher

scores than men on all 22 items and all factors. Mean subscale scores were first calculated for both men and women using Fairburn's original subscales of Restraint (male: $M = 1.20$, $SD = 1.46$; female: $M = 1.63$, $SD = 1.54$), Eating Concern (male: $M = 0.40$, $SD = 0.77$; female: $M = 0.93$, $SD = 1.19$), Weight Concern (male: $M = 0.99$, $SD = 1.19$; female: $M = 1.94$, $SD = 1.61$), and Shape Concern (male: $M = 1.27$, $SD = 1.35$; female: $M = 2.47$, $SD = 1.70$). Mean comparisons were then conducted using Friborg et al.'s (2013) factor structure. The results of the scalar invariance analyses suggest that scores on Items 6 and 10 represent different levels of symptoms in men and women. Both of these items are on Factor 3. Thus, mean comparisons were not carried out for Items 6 and 10 or Factor 3. Due to multiple comparisons, a Bonferroni correction to the alpha level was applied. Since 23 comparisons were conducted, the alpha level used was $.05/23 = .0022$.

As can be seen in Table 5, women had higher scores on 19 of the 20 items that were scalar invariant and all three of the observed factors that were scalar invariant. The only item on which men and women did not differ significantly was the item, "Have you *tried* to follow definite rules regarding your eating (e.g., a calorie limit) in order to influence your shape or weight (whether you have succeeded)?" A comparison of the observed means of the factor scores found that women had higher scores than men on all three factors that did not include the items that lacked scalar invariance. A comparison of latent means suggests that women had higher latent factor scores than men on Factor 1 ($d = 0.528$, $z = 4.059$, $p < .001$), Factor 2 ($d = 0.463$, $z = 5.391$, $p < .001$), and Factor 4 ($d = 0.221$, $z = 7.179$, $p < .001$). To compare scores for latent Factor 3, we allowed the intercepts to differ between men and women on the items that were not scalar invariant (Items 6 and 10). Like the other factors, women had higher scores on latent Factor 3 than did men ($d = 1.106$, $z = 10.584$, $p < .001$).

Discussion

The present study used CFA to examine the factor structure of the EDE-Q in a sample of young adult college students. Analyses compared 12 models of the EDE-Q factor structure supported by previous research. Supplemental analyses examined the goodness-of-fit for nine reduced-item models. The best-fitting model including all subscale items was further assessed for evidence of measurement invariance. Consistent with the great majority of previous research on this topic, support was not found for the original, theoretically derived factor structure. Among those models compared in the CFA, the greatest support was found for Friborg et al.'s (2013) four-factor model. The higher order model was found to fit significantly worse than the first-order model, indicating that use of the global score could be problematic. This finding suggests that eating pathology, as

measured with the EDE-Q, may be multidimensional and not easily summed into a single homogenous score. Thus, creating a global score may obscure differential relations between facets of eating disorder pathology and other constructs in their nomological network. For example, if Factor 1 is negatively correlated with a construct, and Factor 2 is positively correlated with a construct, then a global score may show no correlations with the construct. This result, however, should be interpreted with caution because it is based on a significant difference found for a chi-square difference test. Chi-square difference tests have been shown to be very sensitive to small differences in fit between models (Cheung & Rensvold, 2002). Results from this test may therefore indicate that a more parsimonious model (in this case, the higher order model) fits significantly worse than another model (in this case, the first-order model), when in reality there is no difference in model fit.

Friborg et al.'s (2013) model comprises four factors corresponding to themes of (a) dietary restraint, (b) preoccupation and restriction, (c) weight and shape concern, and (d) eating shame. Factor 1 includes three items from the original Restraint subscale. Factor 2 includes the shape/weight preoccupation item as well as two Eating Concern items and two Restraint items. Factor 3 consists of all items from the Weight Concern and Shape Concern subscales except for the shape/weight preoccupation item and Factor 4 contains three items from the Eating Concern subscale. Peterson's model shares some similarities with the Friborg model (Table S1, see online at <http://journals.sagepub.com/home/asm/>). Most notably, the item loadings for Factor 3 of both the Peterson and Friborg models are consistent with majority of prior factor analysis research in demonstrating that Shape and Weight concern are not conceptually distinct constructs (e.g., Hilbert et al., 2007; White et al., 2014). For this reason, the Weight Concern and Shape Concern subscales of the EDE-Q are often combined into a single, Weight/Shape Concerns Subscale for research purposes (e.g., Mond et al., 2007; van Zutven, Mond, Latner, & Rodgers, 2015).

The items on Factors 1 and 4 suggest partial support for the themes of Fairburn's theoretical model. Factor 1 of both models contains items related to dietary restraint (extreme attempts to limit food intake), while Factor 4 contains a subset of Eating Concern items (eating in secret, guilt after eating, and social eating). This grouping of items from the Eating Concern scale, however, highlights a more specific theme of shame around eating. Shame has been identified by some researchers as a key factor associated with the beliefs and behaviors that may maintain disordered eating (Goss & Allan, 2009). Factor 2 of Friborg et al.'s (2013) model departs the most significantly from Fairburn's original structure. The items included here pertain to dietary restriction (true physiological undereating, not just intention), desire for an empty stomach, fear of loss of control

over eating, and a debilitating preoccupation with eating, shape, and weight. In clinical interpretation, these items may be seen as relating to antecedents and outcomes of starvation. The fear of losing control over one's eating and the desire for an empty stomach may drive efforts to avoid eating and thereby lead to starvation. Starvation, in turn, exacerbates preoccupation with food (Keys, Brožek, Henschel, Mickelsen, & Taylor, 1950).

It may be suggested that dietary restraint and dietary restriction (and the resultant starvation) comprise two distinct features of disordered eating. For example, though dietary restraint is associated with both anorexia and bulimia nervosa, dietary restriction is a diagnostic indicator only for anorexia (American Psychiatric Association, 2013). Thus, an individual with bulimia nervosa may exhibit a pattern of habitual dietary restraint, but may never engage in chronic dietary restriction. This distinction is supported by research from Stice, Fisher, and Lowe (2004) demonstrating that the EDE-Q Restraint subscale did not significantly correlate with actual caloric intake and therefore may not be a valid measure of dietary restriction. The items in Factor 2 may therefore reflect an important and distinct constellation of eating pathology symptoms. Interestingly, while dietary restraint and restriction are not treated as distinct in the EDE-Q, they are distinguished in Fairburn's (2008a) eating disorder treatment guide.

Post hoc analyses using behavioral frequency items as external validators supported the relationship between Friborg et al.'s (2013) factors and behavioral indicators of disordered eating. Of note, Factor 4 (eating shame) was positively associated with all three behavioral variables (binge eating, purging, and excessive exercise), suggesting that this may be an important indicator of disordered eating behavior. Factor 2 (preoccupation and dietary restriction) was not associated with any of these indicators. This is not surprising, however, given that all of the behavioral indicators reflect compensatory behaviors, rather than restriction behaviors (e.g., fasting). Further research should assess this model using additional external validators, such as measures of distress and disability.

In addition to Friborg et al.'s (2013) full model, several of the abbreviated models considered in the current study fit the data well, according to Hu and Bentler's (1998) criteria for goodness-of-fit. The abbreviated models include different data than the full models, which precludes any direct comparison of the fit. As a result, the current research cannot inform a debate about whether items should be removed to increase model fit. Nevertheless, the finding that these models fit that data well is consistent with previous research (e.g., Darcy et al., 2013; Grilo et al., 2013; Hrabosky et al., 2008; Parker et al., 2015, 2016).

Unlike in the two previous studies of sex-related measurement invariance (Grilo et al., 2015; Penelo et al., 2013), the EDE-Q was not found to be scalar invariant, suggesting

that the same scores on the EDE-Q across sexes may represent different levels of eating pathology. Specifically, Item 6 (flat stomach) and Item 10 (fear of weight gain) on Factor 3 (dietary restraint) were found to vary by sex. It is possible that the sex differences on these items reflect differences in ideal body type for men versus women. Research suggests that most young men would like to be lean and muscular rather than lean per se (Labre, 2005; Murray, Griffiths, & Mond, 2016). Among men with eating pathology, many believe their muscles are too small and strive to "bulk up" muscle while maintaining low body fat (Harvey & Robinson, 2003; Murray et al., 2016). For this reason, low weight may not be a goal of some males engaging in disordered eating behavior. Similarly, men with eating disturbances may not desire a "flat stomach," but instead may strive for the "six pack abs" commonly associated with the ideal male body type (Harvey & Robinson, 2003; Murray et al., 2016).

The current findings relating to measurement invariance of the EDE-Q by sex reflect a broader concern within the field regarding the "female-centric" nature of diagnostic criteria for eating disorders and consequently, instruments such as the EDE-Q that are designed to assess these criteria (Mitchison & Mond, 2015). In this regard, however, it is notable that a lack of measurement invariance was apparent only in the case of the two aforementioned items and that comparison between sexes on three of the four factors was, in fact, supported. Methodologists have suggested that a lack of measurement invariance may not be a problem if fewer than 20% of the items are responsible for a lack of measurement invariance (Byrne et al., 1989). In the current research, the two items that were responsible for the lack of scalar invariance represent only 9.1% of the items. Thus, the lack of scalar invariance may not be a problem when comparing scores.

Furthermore, while one interpretation for a lack of scalar invariance suggests that the same score represents different levels of eating pathology in men and women, it also is possible that the same score indicate something equivalent between groups. Ideally, these alternative interpretations would be tested by comparing correlations between EDE-Q scores and scores on an appropriate external criterion between males and females. While this was not possible in the current study, it may be noted that findings from recent, population-based studies suggest that, in terms of associations with measures of distress and impairment in role functioning, items of the EDE-Q do indeed function similarly in males and females (Bentley, Gratwick-Sarll, Harrison, & Mond, 2015; Bentley, Mond, & Rodgers, 2014). These findings suggest that the EDE-Q may in fact be appropriate for use in both males and females.

Similarly, one solution to the lack of scalar invariance could be to remove the two problematic items and compare scores on the rest of the scales. However, this may remove

important aspects of the construct of interest. Since the scale had metric invariance, the lack of scalar invariance may only be an issue for comparing scores between men and women. An alternative approach may be to model the EDE-Q scores as a latent variable and compare latent means while allowing the intercepts for Items 6 and 10 to differ between groups.

There are limitations to the current research. One limitation relates to data on model validity. In the current sample, no measures were collected to provide evidence of construct validity for the proposed model. If additional measures were included, they could have been used as another test to determine how well the various factor models describe eating pathology. For example, they could be used as a test to compare the construct validity of the shortened scales, many of which provide good fit to the data, with the full versions of the scale. There are also some limitations related to our sample. The models included in our analyses were developed and tested using samples that varied in age, nationality, and clinical status. Though different in mean age and nationality, both the present study and Friborg et al. (2013) utilized a nonclinical community sample in model testing. This similarity in samples may have contributed to the fit of their model to our sample.

Furthermore, the findings presented here are based on a convenience sample of college students. As our sample is educated and predominantly young, healthy, and normal weight, our results may not generalize to general population samples of young men and women, to clinical samples of individuals with different eating disorder diagnoses, or to individuals of other age groups. To improve the generalizability of our findings, it is important that future research test this model, as well as its measurement invariance by diagnostic status, in a clinical sample. Additionally, the ethnic diversity of our sample may be viewed as both a strength and weakness. While our sample provides much needed research on understudied Asian minorities, it may also reduce the generalizability of our findings among different demographic samples from the mainland United States and other countries. Research addressing measurement invariance between Asian American and non-Asian participants would be of interest in this regard.

Furthermore, nearly 70% of the participants in our sample were female. Though the number of male participants in our sample provided sufficient statistical power to evaluate measurement invariance, future research should further explore issues of sex-related measurement invariance with larger samples of male participants. In the current research, we used multigroup CFA to examine the measurement invariance of the EDE-Q. Future research could use exploratory structural equation modeling to examine the measure invariance of the scale, as has been done with other scales in previous research (Asparouhov & Muthén, 2009). Such an approach could

lead to a different factor structure being identified as the best-fitting model, and could identify different items that are valuable to understanding eating pathology in males and females.

In sum, the current research presents a clinically meaningful and psychometrically supported factor structure for the EDE-Q. This model improves on the empirical support of the original model, while maintaining similar descriptive utility by retaining all the original items. Our findings have several implications for researchers and clinicians using the EDE-Q. Overall, these results expand on a robust body of research suggesting that users should consider moving away from the rationally derived factor structure of the current EDE-Q. More specifically, this model reinforces previous research indicating that shape and weight concern are not factorially distinct constructs and should not be interpreted separately. The current model also highlights the importance of eating shame and dietary restriction in describing disordered eating. Employing the suggested factor structure will allow clinicians and researchers to examine the role of these factors in eating behavior. Additionally, these findings provide empirical support for the appropriateness of comparisons by sex on most factors, though caution should be exercised when making comparisons by sex on Factor 3. To date only a few studies have examined measurement invariance on the EDE-Q and further research is needed in this area. By taking these considerations into account, clinicians and researchers may improve their interpretations of this measure and their understanding of the structure and nature of the eating disturbances it assesses.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplementary Material

Supplementary material for this article is available online.

Note

1. Models were also specified using weighted least squares mean and variance estimation. Results were not found to differ from those with the unweighted least squares mean and variance estimator.

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*, 305-309.

- 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*, 143-151. doi:10.1016/j.eatbeh.2011.01.005
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*, 397-438. doi:10.1080/10705510903008204
- Barnes, J., Prescott, T., & Muncer, S. (2012). Confirmatory factor analysis for the Eating Disorder Examination Questionnaire: Evidence supporting a three-factor model. *Eating Behaviors, 13*, 379-381. doi:10.1016/j.eatbeh.2012.05.001
- Becker, A. E., Thomas, J. J., Bainivualiku, A., Richards, L., Navara, K., Roberts, A. L., . . . Striegel-Moore, R. H. (2010). Validity and reliability of a Fijian translation and adaptation of the Eating Disorder Examination Questionnaire. *International Journal of Eating Disorders, 43*, 173-178.
- Bentley, C., Gratwick-Sarll, K., Harrison, C., & Mond, J. (2015). Sex differences in psychosocial impairment associated with eating disorder features in adolescents: A school-based study. *International Journal of Eating Disorders, 48*, 633-640. doi:10.1002/eat.22396
- Bentley, C., Mond, J., & Rodgers, B. (2014). Sex differences in psychosocial impairment associated with eating-disordered behavior: What if there aren't any? *Eating Behaviors, 15*, 609-614. doi:10.1016/j.eatbeh.2014.08.015
- 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. doi:10.1002/eat.20931
- Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456-466.
- Calugi, S., Milanese, C., Sartirana, M., El Ghoch, M., Sartori, F., Geccherle, E., . . . Dalle Grave, R. (2016). The Eating Disorder Examination Questionnaire: Reliability and validity of the Italian version. *Eating and Weight Disorders, 22*, 509-514. doi:10.1007/s40519-016-0276-6
- Carrard, I., My Lien Rebetz, M., Mobbs, O., & Van der Linden, M. (2015). Factor structure of a French version of the Eating Disorder Examination-Questionnaire among women with and without binge eating disorder symptoms. *Eating and Weight Disorders, 20*, 137-144. doi:10.1007/s40519-014-0148-x
- Cash, T. F. (Ed.). (2012). *Encyclopedia of body image and human appearance*. Waltham, MA: Academic Press.
- Chan, C. W., & Leung, S. F. (2015). Validation of the Eating Disorder Examination Questionnaire: An online version. *Journal of Human Nutrition and Dietetics, 28*, 659-665. doi:10.1111/jhn.12275
- Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology, 95*, 1005-1018. doi:10.1037/a0013193
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*, 233-255. doi:10.1207/S15328007SEM0902_5
- Cooper, Z., Cooper, P. J., & Fairburn, C. G. (1989). The validity of the Eating Disorder Examination and its subscales. *British Journal of Psychiatry, 154*, 807-812.
- Cooper, Z., & Fairburn, C. G. (1987). The Eating Disorder Examination: A semi-structured interview for the assessment of the specific psychopathology of eating disorders. *International Journal of Eating Disorders, 6*, 1-8.
- Darcy, A. M., Hardy, K. K., Crosby, R. D., Lock, J., & Peebles, R. (2013). Factor structure of the Eating Disorder Examination Questionnaire (EDE-Q) in male and female college athletes. *Body Image, 10*, 399-405. doi:10.1016/j.bodyim.2013.01.008
- Fairburn, C. G. (2008a). *Cognitive behavior therapy and eating disorders*. New York, NY: Guilford Press.
- Fairburn, C. G. (2008b). Eating Disorder Examination Questionnaire (EDE-Q 6.0). In C. G. Fairburn (Ed.), *Cognitive behavior therapy and eating disorders* (pp. 309-313). New York, NY: Guilford Press.
- Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders, 16*, 363-370.
- Fairburn, C. G., & Cooper, Z. (1993). The Eating Disorder Examination (12th edition). In C. G. Fairburn & G. T. Wilson (Eds.), *Binge eating: Nature, assessment, and treatment* (pp. 317-360). New York, NY: Guilford Press.
- Franko, D.L., Jenkins, A., Roehrig, J.P., Luce, K.H., Crowther, J.H., & Rodgers, R.F. (2012). Psychometric properties of measures of eating disorder risk in Latina college women. *International Journal of Eating Disorders, 45*(4), 592-596. doi:10.1002/eat.20979
- 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*, 195-203. doi:10.1002/mpr.1389
- Gideon, S., Hawkes, N., Mond, J.M., Saunders, R., Tchanturia, K., & Serpell, L. (2016). Development and psychometric validation of the EDE-QS, a 12 item short form of the Eating Disorder Examination Questionnaire (EDE-Q). *PLoS One, 11*(5), e0152744. doi: 10.1371/journal.pone.0152744
- Giozazolias, T., Tsaousis, I., & Vallianatou, C. (2013). The factor structure and psychometric properties of the Greek version of the Eating Disorders Examination Questionnaire (EDE-Q). *European Journal of Psychological Assessment, 29*, 189-196. doi:10.1027/1015-5759/a000138
- Goss, K., & Allan, S. (2009). Shame, pride and eating disorders. *Clinical Psychology & Psychotherapy, 16*, 303-316. doi:10.1002/cpp.627
- Grilo, C. M., Crosby, R. D., Peterson, C. B., Masheb, R. M., White, M. A., Crow, S. J., . . . Mitchell, J. E. (2010). Factor structure of the Eating Disorder Examination interview in patients with binge-eating disorder. *Obesity, 18*, 977-981. doi:10.1038/oby.2009.321
- 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*, 657-662. doi: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*, 284-289. doi:10.1002/eat.22358
- Guest, T. (2000). Using the Eating Disorder Examination in the assessment of bulimia and anorexia: Issues of reliability and validity. *Social Work in Health Care, 31*(4), 71-83. doi:10.1300/J010v31n04_05
- Harvey, J. A., & Robinson, J. D. (2003). Eating disorders in men: Current considerations. *Journal of Clinical Psychology in Medical Settings, 10*, 297-306. doi:10.1023/A:1026357505747
- Heene, M., Hilbert, S., Freudenthaler, H. H., & Bühner, M. (2012). Sensitivity of SEM fit indexes with respect to violations of uncorrelated errors. *Structural Equation Modeling: A Multidisciplinary Journal, 19*, 36-50. doi:10.1080/10705511.2012.634710
- Hilbert, A., de Zwaan, M., & Braehler, E. (2012). How frequent are eating disturbances in the population? Norms of the eating disorder examination-questionnaire. *PLoS ONE, 7*, e29125. doi:10.1371/journal.pone.0029125
- Hilbert, A., Tuschen-Caffier, B., Karwautz, A., Niederhofer, H., & Munsch, S. (2007). Eating Disorder Examination-Questionnaire. *Diagnostica, 53*, 144-154. doi:10.1026/0012-1924.53.3.144
- 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. doi:10.1038/oby.2008.3
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424-453. doi:10.1037/1082-989X.3.4.424
- Hudson, J. I., Hiripi, E., Pope, H. G., & Kessler, R. C. (2007). The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biological Psychiatry, 61*, 348-358. doi:10.1016/j.biopsych.2006.03.040
- Keys, A., Brožek, J., Henschel, A., Mickelsen, O., & Taylor, H. L. (1950). *The biology of human starvation* (Vol. 1-2). Minneapolis: University of Minnesota Press.
- 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*, 613-616. doi:10.1002/eat.22487
- Labre, M. (2005). Burn fat, build muscle: A content analysis of men's health and men's fitness. *International Journal of Men's Health, 4*, 187-200. doi:10.3149/jmh.0402.187
- Luce, K. H., & Crowther, J. H. (1999). The reliability of the Eating Disorder Examination-Self-Report Questionnaire Version (EDE-Q). *International Journal of Eating Disorders, 25*, 349-351.
- Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin, 97*, 562-582. doi:10.1037/0033-2909.97.3.562
- McDonald, R. P. (1989). An index of goodness-of-fit based on non-centrality. *Journal of Classification, 6*, 97-103. doi:10.1007/BF01908590
- McGartland Rubio, D., Berg-Weger, M., & Tebb, S. S. (2001). Using structural equation modeling to test for multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal, 8*, 613-626. doi:10.1207/S15328007SEM0804_06
- Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*, 568-592. doi:10.1037/0021-9010.93.3.568
- Mitchison, D., Hay, P., Slewa-Younan, S., & Mond, J. (2014). The changing demographic profile of eating disorder behaviors in the community. *BMC Public Health, 14*, 943. doi:10.1186/1471-2458-14-943
- Mitchison, D., & Mond, J. (2015). Epidemiology of eating disorders, eating disordered behaviour, and body image disturbance in males: A narrative review. *Journal of Eating Disorders, 3*, 20. doi:10.1186/s40337-015-0058-y
- Mond, J. M., Hay, P. J., Rodgers, B., & Owen, C. (2006). Eating Disorder Examination Questionnaire (EDE-Q): Norms for young adult women. *Behaviour Research and Therapy, 44*, 53-62. doi:10.1016/j.brat.2004.12.003
- Mond, J. M., Hay, P. J., Rodgers, B., Owen, C., & Beumont, P. J. V. (2004a). Temporal stability of the Eating Disorder Examination Questionnaire. *International Journal of Eating Disorders, 36*, 195-203. doi:10.1002/eat.20017
- Mond, J. M., Hay, P. J., Rodgers, B., Owen, C., & Beumont, P. J. V. (2004b). Validity of the Eating Disorder Examination Questionnaire (EDE-Q) in screening for eating disorders in community samples. *Behaviour Research and Therapy, 42*, 551-567. doi:10.1016/S0005-7967(03)00161-X
- Mond, J. M., Rodgers, B., Hay, P. J., Darby, A., Owen, C., Baune, B. T., & Kennedy, R. L. (2007). Obesity and impairment in psychosocial functioning in women: The mediating role of eating disorder features. *Obesity, 15*, 2769-2779. doi:10.1038/oby.2007.329
- Murray, S. B., Griffiths, S., & Mond, J. M. (2016). Evolving eating disorder psychopathology: Conceptualising muscularity-oriented disordered eating. *British Journal of Psychiatry, 208*, 414-415. doi:10.1192/bjp.bp.115.168427
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide: Seventh edition*. Los Angeles, CA: Muthén & Muthén.
- Parker, K., Mitchell, S., O'Brien, P., & Brennan, L. (2015). Psychometric evaluation of disordered eating measures in bariatric surgery patients. *Eating Behaviors, 19*, 39-48. doi:10.1016/j.eatbeh.2015.05.007
- 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. doi:10.1007/s11695-015-1780-x
- Penelo, E., Negrete, A., Portell, M., & Raich, R. M. (2013). Psychometric properties of the Eating Disorder Examination Questionnaire (EDE-Q) and norms for rural and urban adolescent males and females in Mexico. *PLoS ONE, 8*, e83245. doi:10.1371/journal.pone.0083245
- Peterson, C. B., Crosby, R. D., Wonderlich, S. A., Joiner, T., Crow, S. J., Mitchell, J. E., . . . 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. doi:10.1002/eat.20373

- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*, 354-373. doi:10.1037/a0029315
- Satorra, A., & Bentler, P.M. (2001). A scale difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507-514. doi: 10.1007/BF02296192
- Sörbom, D. (1989). Model modification. *Psychometrika, 54*, 371-384. doi:10.1007/BF02294623
- Stanford, S. C., & Lemberg, R. (2012). Measuring eating disorders in men: Development of the eating disorder assessment for men (EDAM). *Eating Disorders, 20*, 427-436. doi:10.1080/10640266.2012.715522
- Stice, E., Fisher, M., & Lowe, M. R. (2004). Are dietary restraint scales valid measures of acute dietary restriction? Unobtrusive observational data suggest not. *Psychological Assessment, 16*, 51-59. doi:10.1037/1040-3590.16.1.51
- van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology, 4*, 770. doi:10.3389/fpsyg.2013.00770
- van Zutven, K., Mond, J., Latner, J., & Rodgers, B. (2015). Obesity and psychosocial impairment: Mediating roles of health status, weight/shape concerns and binge eating in a community sample of women and men. *International Journal of Obesity, 39*, 346-352. doi:10.1038/ijo.2014.100
- Vartanian, L. R. (2012). Self-discrepancy theory and body image. In T. F. Cash (Ed.), *Encyclopedia of body image and human appearance* (pp. 711-717). Waltham, MA: Academic Press.
- Villarroel, A.M., Penelo, E., Portell, M., & Raich, R.M. (2011). Screening for eating disorders in undergraduate women: Norms and validity of the Spanish version of the Eating Disorder Examination Questionnaire (EDE-Q). *Journal of Psychopathology and Behavioral Assessment, 33*, 121-128. doi: 10.1007/s10862-009-9177-6
- Wade, T. D., Byrne, S., & Bryant-Waugh, R. (2008). The eating disorder examination: Norms and construct validity with young and middle adolescent girls. *International Journal of Eating Disorders, 41*, 551-558. doi:10.1002/eat.20526
- White, H. J., Haycraft, E., Goodwin, H., & Meyer, C. (2014). Eating disorder examination questionnaire: Factor structure for adolescent girls and boys. *International Journal of Eating Disorders, 47*, 99-104. doi:10.1002/eat.22199