



Published in final edited form as:

Int J Eat Disord. 2012 April ; 45(3): 428–438. doi:10.1002/eat.20931.

Psychometric Evaluation of the Eating Disorder Examination and Eating Disorder Examination-Questionnaire: A Systematic Review of the Literature

Kelly C. Berg, PhD^{1,2,*}, Carol B. Peterson, PhD², Patricia Frazier, PhD¹, and Scott J. Crow, MD²

¹Department of Psychology, University of Minnesota, Minneapolis, Minnesota

²Department of Psychiatry, University of Minnesota, Minneapolis, Minnesota

Abstract

Objective—The purpose of this study was to systematically review the reliability of scores on the Eating Disorder Examination (EDE) and the Eating Disorder Examination-Questionnaire (EDE-Q) and to examine the validity of their use as measures of eating disorder symptoms.

Method—Articles describing the psychometric properties of the EDE and EDE-Q were identified in a systematic search of major computer databases and a review of reference lists. Articles were selected based on a priori inclusion and exclusion criteria.

Results—Fifteen studies were identified that examined the psychometrics of the EDE, whereas 10 studies were found that examined the psychometrics of the EDE-Q.

Discussion—Both instruments demonstrated reliability of scores. There is evidence that scores on the EDE and EDE-Q correlate with scores on measures of similar constructs and support for using the instruments to distinguish between cases and non-cases. Additional research is needed to broaden the generalizability of the findings.

Keywords

eating disorder examination; eating disorder examination-questionnaire; reliability; validity; psychometrics

Introduction

The interview (EDE)^{1,2} and questionnaire (EDE-Q)^{3,4} versions of the Eating Disorder Examination are widely considered the preeminent eating disorder assessments.⁵ Both instruments include four sub-scales related to the cognitive features of eating disorders: Restraint, Eating Concern, Shape Concern, and Weight Concern. They also include items that assess specific behavioral symptoms, such as the frequency of binge eating, self-induced vomiting, laxative misuse, diuretic misuse, and excessive exercise. Researchers and clinicians alike use the EDE and EDE-Q to obtain descriptive information regarding eating disorder (ED) symptoms and to make eating disorder diagnoses. Additionally, its status as the gold standard of ED assessment has given the EDE the weighty responsibility of serving to validate other assessments.^{6,7}

Despite its prominence, relatively few studies have examined the reliability of EDE and EDE-Q scores or whether scores on the two instruments are valid measures of ED symptoms. Additionally, a review of the psychometric properties of these instruments has never been published. Thus, the purpose of this article is to systematically review research on the psychometric properties of the EDE and EDE-Q, evaluate the support for the psychometrics of the two instruments, and provide recommendations for future research.

Method

A literature search was conducted for studies that assessed the psychometric properties of the EDE and EDE-Q using three major computer databases (i.e., MED-LINE, PsycINFO, PubMed) and by reviewing reference lists from published journal articles and books. Search terms included “Eating Disorder Examination” and “Eating Disorder Examination-Questionnaire.” Studies were included if the purpose of the study was to examine one or more of the following psychometric properties of the EDE or EDE-Q: test–retest reliability, inter-rater reliability, internal consistency, content validity, criterion-oriented validity, or construct validity. The literature search was inclusive of studies that assessed the psychometric properties of the four cognitive subscales (Restraint, Eating Concern, Shape Concern, and Weight Concern), Objective Bulimic Episodes (OBEs), and Subjective Bulimic Episodes (SBEs), self-induced vomiting, laxative misuse, diuretic misuse, or excessive exercise. Studies published in languages other than English and those examining the psychometric properties of translated^{8,9} or child¹⁰ versions of the EDE or EDE-Q were excluded.

Results

Reliability of EDE Scores

Test–retest reliability—Two research groups have examined the test–retest reliability of EDE scores^{11,12} (see Table 1). The first study assessed short-term (2–7 days) test–retest reliability in 20 women with a variety of eating disorders¹² and the second reported the test–retest reliability of EDE scores over a relatively longer period of time (6–14 days) in 18 women with Binge Eating Disorder (BED).¹¹ In both studies, test–retest reliability coefficients for the four subscale scores ranged from 0.50 to 0.88. With the exception of SBEs, scores on the behavioral frequency items also demonstrated test–retest reliability, with correlations ranging from 0.70 to 0.97.

Inter-rater reliability—Because the EDE is a semi-structured interview, it is important to examine whether raters make similar ratings. One study examined the inter-rater reliability of scores on each individual EDE item.¹⁴ Three raters each assessed 12 different participants, nine of whom met criteria for Bulimia Nervosa (BN) and three of whom had no eating disorder. Of the 62 total items examined (some of which have since been removed from the EDE), scores on 27 items had perfect inter-rater reliability and scores on only three items had inter-rater reliability coefficients below 0.90, two of which remain in the EDE (social eating and body composition).

Three studies^{11,12,15} have examined the inter-rater reliability of scores on the four subscales of the EDE and the behavior frequency items^a. The first used a sample of 106 undergraduate women,¹⁵ the second sampled 20 women with a variety of eating disorders,¹² and, in the third, participants were 18 adult women with BED.¹¹ Across the three studies, the inter-rater

^aSeveral studies have reported the inter-rater reliability coefficients for the EDE within the context of other studies (e.g., Ref. 16); however, these data are not consistently reported. This review only includes studies whose purpose was to examine the inter-rater reliability of the EDE.

reliability coefficients of scores on the four subscales ranged from 0.65 to 0.99. The inter-rater reliability coefficients for scores on the behavior frequency items^b ranged from 0.91 to 1.0 (see Table 2).

Internal consistency—Four studies^c have examined the internal consistency of the four subscales of the EDE in six samples.^{17–20} The first study sampled an ED population that included 47 women with Anorexia Nervosa (AN), 53 women with BN, and 42 controls.¹⁹ Participants in the second study were 116 adult women with various eating disorders.¹⁷ Participants in the third study were 688 adults seeking treatment for BED.²⁰ Finally, the fourth study,¹⁸ examined the internal consistency of the EDE sub-scales in three samples: a female ED sample including 24 participants with AN, 67 with BN, and 67 with Eating Disorder Not Otherwise Specified (EDNOS); 317 women from a community sample; and 170 women seeking treatment for overweight or obesity. The internal consistency coefficients of the subscales ranged from 0.58 to 0.78 for the Restraint subscale, 0.44 to 0.78 for the Eating Concern subscale, 0.68 to 0.85 for the Shape Concern subscale, and 0.51 to 0.76 for the Weight Concern subscale (see Table 3).

Long term recall—The EDE asks individuals to recall symptoms that occurred up to 6 months before the interview, but there are little data to suggest that individuals are able to recall these symptoms accurately. Although research supports the test–retest reliability or repeatability of EDE scores, this does not demonstrate that individuals accurately recall past symptoms. Two studies^{22,24} have assessed the long-term recall accuracy of ED symptoms using the EDE. In both studies, participants with a variety of eating disorders completed a first EDE at time 1 and a second EDE at a follow-up assessment. During the second EDE, they were asked to recall symptoms from time 1 rather than current symptoms. In the first study,²² the second EDE was conducted either 6 or 12 months after the first EDE, whereas in the second study²⁴ the second EDE was conducted 5 to 30 months after the first EDE. There were significant correlations between the subscale scores at baseline and recall, with correlations ranging from 0.63 to 0.88.²⁴ With the exception of SBEs in the first study,²² there were also significant correlations between the frequencies of binge eating and compensatory behaviors reported at baseline and recall (range of $r_s = 0.65$ to 0.97). The researchers from the first study²² also compared diagnoses at time 1 and at recall. The agreement rate (65%) and kappa coefficients (.42 to 0.50) were lower for narrow diagnoses (e.g., AN) than the agreement rates (78 to 86%) and kappa coefficients (.56 to 0.65) for broad diagnoses (e.g., full-threshold eating disorder).

Reliability of EDE-Q Scores

Test-retest reliability—The test–retest reliability of EDE-Q scores over 1 to 14 days has been examined by two groups of researchers,^{7,13} first in a community sample of 139 female undergraduate students¹³ and second in a sample of 86 men and women seeking treatment for BED.⁷ Across studies, the test–retest correlations ranged from 0.66 to 0.94 for scores on the four subscales and from 0.51 to 0.92 for the behavior frequency items (see Table 1). Test–retest reliability correlations were also significant for scores on the individual items that are used to create the four subscales⁷ with correlations ranging from 0.40 (fear of weight gain) to 0.78 (importance of shape, reaction to prescribed weighing). One study analyzed the test–retest correlations across different time lags⁷ and found that time lag had

^bRosen et al., (1990) did not differentiate between OBEs and SBEs in their analyses. Although the authors do not provide the explicit criteria used to define “binge eating,” it is assumed that when the term “binge eating” is used, it is meant to describe what should be termed as OBEs. This assumption will be applied to all other studies cited in this paper that analyzed frequency of binge eating without distinguishing between OBEs and SBEs.

^cThe internal consistency of the EDE is rarely reported in the literature; thus, this review only includes studies whose purpose was to examine the internal consistency of the EDE.

little impact on the test–retest correlations for the EDE subscale scores or scores on the items measuring frequency of OBEs. However, for scores on the items assessing the frequency of SBEs and Objective Overeating Episodes (OOEs) test–retest reliability coefficients decreased as the time lag increased.

Temporal stability—The temporal stability of EDE-Q scores over 5–14 months has been examined in two published studies^{21,25}; first in a community sample of 196 Australian women²¹ and later in a community sample of 70 black and 156 white American women.²⁵ The temporal stability of the subscale scores were comparable for black and white women²⁵ and to the short-term test–retest correlations for EDE-Q scores found by Luce and Crowther (1999) and Reas et al. (2006).^{21,25} Scores on the items measuring frequency of OBEs, SBEs, and excessive exercise also demonstrated temporal stability; however, these correlations were weaker than the correlations for the EDE-Q subscale scores or the short-term test–retest correlations²¹ (see Table 4). Additionally, one study examined the temporal stability of scores on the individual items of the EDE-Q and found that the correlations between individual items rated at time 1 and time 2 were all significant, ranging from 0.42 (“eating in secret”) to 0.69 (“feelings of fatness”).²¹

Internal consistency—Four studies have assessed the internal consistency of the EDE-Q sub-scales.^{13,21,23,25} These studies included a community sample of 203 undergraduate women at time 1 and 139 (68.5%) of the women at time 2,¹³ a community sample of 208 adult women,²¹ a community sample of 70 black and 156 white women,²⁵ and 203 adult women with BN.²³ The four subscales demonstrated acceptable internal consistency in all four studies, with alphas ranging from 0.70 to 0.93 (see Table 3). One study also calculated the item-total correlations for the EDE-Q and found correlations ranging from 0.33 (“avoidance of eating,” “eating in secret”) to 0.76 (“dissatisfaction with weight,” “dissatisfaction with shape”).²¹

Validity of the EDE for the Assessment of ED Symptoms

Criterion-oriented validity: Ability to detect group differences—One method for testing criterion-oriented validity is to determine whether the instrument predicts expected group differences.^{26,27} Four studies have examined the ability of EDE scores to discriminate between ED populations and control groups.^{15,19,28,29} In the first of these studies,¹⁹ the EDE scores of 47 women with AN, 53 women with BN, and 42 control women were compared. Two studies have examined the ability of EDE scores to discriminate between women with BN and control women who score high on a measure of restraint.^{15,29} The final study compared 105 adult women with BED to a group of 42 normal-weight and 15 overweight control women.²⁸ Data from these studies show that there were large effect sizes for differences in scores on the four subscales between the following groups: AN and control women, BN and control women, BN and restricting control women, BED and normal-weight control women, and BED and overweight control women (range of Cohen’s d s = 0.97 to 6.68). The only exceptions were a moderate difference between the BN and restricting control women on the Restraint subscale and a small difference between the BED and overweight control women on the Restraint subscale. Additionally, the EDE demonstrated the ability to differentiate women with AN and BN on items measuring OBE frequency¹⁹ (see Table 5)^d.

Construct validity: Convergence with measures of similar constructs—One method of testing construct validity is to determine whether two different measures of a

^dThe results from Rosen et al. (1990) are not included in the table as the authors only described the results of the group differences comparisons within text and did not report specific statistics.

construct converge.²⁶ Three studies^e have examined whether scores on the four subscales of the EDE correlate with measures of similar constructs. The first used a sample of 106 undergraduate women,¹⁵ the second used a sample of 82 women seeking treatment for BN,³² and the third used a sample of 66 obese adults with subthreshold BED.³³ In the first two studies,^{15,32} scores on all four subscales of the EDE correlated with measures of similar constructs (see Table 6). The third study³³ found significant correlations between scores on the following individual items on the EDE and scores on the associated items of the Questionnaire on Eating and Weight Patterns-Revised (QEWP-R)³⁴: OBE frequency (0.33), distress regarding OBEs (0.32), overevaluation of shape (0.36) and weight (0.37), and criticism of self because of shape and weight (0.45).

Construct validity: Convergence with daily food records—Three studies have examined whether scores on the behavioral frequency items of the EDE correlate with the frequency of binge eating and compensatory behaviors reported on daily food records (DFRs).^{15,32,35} In the first of these studies,¹⁵ scores on the behavioral frequency items of the EDE were compared with 7 days of DFRs in a community sample of 106 women. A second study correlated the frequency of binge eating and compensatory behaviors reported on the EDE with the frequency of these behaviors reported on DFRs for both a 7-day and 28-day time period in a sample of 82 women seeking treatment for BN.³² The third study³⁵ examined whether the frequency of binge eating and compensatory behaviors reported on the EDE correlated with the frequency of these behaviors reported on DFRs in 16 women diagnosed with either BN or subthreshold AN binge/purge subtype.

All three studies found significant positive relationships between the frequency of binge episodes reported on the EDE and the DFRs, with correlations ranging from 0.56 to 0.93. Similarly, the correlations between the frequency of compensatory behaviors reported on the EDE and DFRs ranged from 0.62 to 1.00 (see Table 7). Only one study conducted a comparison of means analysis³⁵ and found that participants reported significantly higher rates of binge eating and excessive exercise on the EDE than on DFRs (Cohen's $d = 0.42$ and 0.78 , respectively). There were no significant differences in the frequency of self-induced vomiting, laxative use, or diuretic use reported on the EDE and DFRs with Cohen's d s ranging from 0.13 to 0.24.

Construct validity: Factor structure of the EDE—As stated previously, the EDE is conceptualized as having four subscales: Restraint, Weight Concern, Shape Concern, and Weight Concern and three studies have examined the factor structure of the EDE.^{18,20,36} The factor structure of the EDE has been examined in a sample of 115 obese adults who did not meet criteria for BED,³⁶ in a sample of 688 adults seeking treatment for BED,²⁰ and in a sample of 158 adolescent and adult women with eating disorders, 170 adult women seeking treatment for obesity, and 317 control women.¹⁸ None of these studies replicated the EDE's four-factor model. In the first study, a 2-factor model was the best fit.³⁶ In this model, the first factor was similar to the Restraint subscale and the second appeared to be a combination of the remaining three sub-scales. The second study found a 3-factor model (i.e., “Dietary Restraint,” “Shape/Weight Overevaluation,” and “Body Dissatisfaction”) which was supported by a confirmatory factor analysis. Finally, the third study compared the original four-factor structure of the EDE to three-, two-, and one-factor models and found that a one-factor model (i.e., “Weight and Shape Concern”) was the best fit.

^eAdditional studies have examined the convergent validity of the EDE and other self-report questionnaires (e.g., Ref. 30) as well as the convergent validity of the EDE and other interview-based assessments (e.g., Ref. 31); however the purpose of these studies has been to examine the validity of the other instrument against the EDE. These studies are not reported here as it does not seem suitable to discuss the psychometric properties of the EDE against unvalidated instruments.

Validity of the EDE-Q for the Assessment of ED Symptoms

Criterion-oriented validity: Ability to detect group differences—Four studies have examined whether EDE-Q scores discriminate between eating disorder and control groups.^{37–40} In one study, a structured interview was used to classify 13 women as cases of eating disorders (i.e., non-purging BN, EDNOS) and 182 women as noncases of eating disorders.³⁹ Women diagnosed with eating disorders scored significantly higher on the EDE-Q than the control women. Two additional studies used the EDE-Q to classify participants as cases and noncases of eating disorders.^{38,40} In the first study, obese binge eaters scored significantly higher than obese non-binge eaters on 15 individual items of the EDE-Q⁴⁰. The items that did not discriminate between the two groups were items that reflected dietary restraint and a desire to lose weight. It is worth noting that the entire sample in this study was drawn from a weight loss program; thus, one may not expect differences between groups on these variables. The second study³⁸ found that female adolescents with AN scored significantly higher than controls on the Eating Disorders Inventory (EDI)⁴¹ and all but one subscale of the 12-item version of the Eating Attitudes Test (EAT-12),⁴² with effect sizes ranging from 0.87 to 1.56.

Finally, one study examined the agreement between the EDE-Q and another self-report measure of binge eating, the QEWP-R³⁴, in identifying regular binge eaters among 249 adult bariatric surgery candidates.³⁷ When binge eating was defined as having at least 1 episode of binge eating per week, approximately the same number of participants was classified as binge eaters by the EDE-Q (20.7%) and QEWP-R (23.2%), yet the agreement between those measures was low (Cohen's kappa = 0.26). When binge eating was defined as having at least 2 binge episodes per week, the QEWP-R identified 1.5 times as many binge eaters as did the EDE-Q (13.9 and 8.9%, respectively) and the agreement between the EDE-Q and QEWP-R in identifying twice-weekly binge eaters was almost entirely due to chance (Cohen's kappa = 0.05).

Construct validity: Convergence with assessments of similar constructs—

Only one study has examined whether scores on the four subscales of the EDE-Q correlate with measures of similar constructs.²⁵ Using a community-based sample of 70 Black and 156 White women, the researchers demonstrated that scores on the Restraint subscale of the EDE-Q correlated more strongly to scores on other measures of restraint ($r_s = 0.76$ to 0.79) than to scores on measures of bulimic symptoms ($r_s = 0.40$ to 0.54).

Construct validity: Convergence with daily food records—

Two studies, both using samples of treatment seeking adults with BED, have examined whether the frequencies of OBEs, SBEs, and OOEes reported on the EDE-Q correlate with the frequencies of the same behaviors reported on DFRs.^{6,43} In both studies, there were significant positive correlations between the frequency of OBEs, SBEs, and OOEes reported on the EDE-Q and on the DFRs, with correlations ranging from 0.31 to 0.63. There were no significant differences between the number of OBE episodes reported on the EDE-Q and the DFRs in either study. However, participants reported significantly more SBEs and OOEes on the DFRs than on the EDE-Q (Cohen's $d_s = 0.53$ to 1.13).

Construct validity: Factor structure of the EDE-Q—

Two studies have examined the factor structure of the EDE-Q.^{23,44} The first study²³ conducted an exploratory factor analysis using a sample of 203 adults with full- and sub-threshold BN and the second study⁴⁴ conducted both exploratory and confirmatory factor analyses in a sample of 337 adult obese bariatric surgery candidates. The results of the both studies supported a four-factor model; however, neither replicated the original four subscales of the EDE-Q. Both studies found a factor that resembled the Restraint subscale, a second factor that appeared to

be a combination of items from the Shape Concern and Weight Concern subscales, and a third factor that included only the importance of shape and weight items. In the first study,²³ the remaining factor resembled the Eating Concern subscale; in the second study,⁴⁴ the remaining factor consisted of items assessing overeating or binge eating and appeared to describe general disturbances in eating behavior.

Discussion

Reliability of EDE Scores

Current research provides support for the test–retest reliability of scores on the four subscales, scores on the individual items that assess objective bulimic days and episodes, and scores on the individual items that assess self-induced vomiting days and episodes. However, with the exception of the Restraint subscale, the test–retest reliability correlations weakened as the length of time between testing increased, although the time between testing was not long (i.e., 2–14 days). The data do not support the test–retest reliability for scores on the items that assess SBEs and there are no published data on the test–retest reliability of scores on the EDE items that assess laxative or diuretic misuse. However, it should be noted that only two studies have assessed the test–retest reliability of EDE scores and both had relatively small sample sizes.

Published data support the inter-rater reliability of scores on the individual items of the EDE, scores on the four subscales of the EDE, and scores on the individual items that assess binge eating and self-induced vomiting. The lowest inter-rater reliability coefficients were from the Rosen et al. (1990) study, which was the only one that used a nonclinical sample. No published studies have assessed the inter-rater reliability of scores on the items that assess laxative misuse, diuretic misuse, or excessive exercise. Additionally, it is notable that at least two of the studies calculated inter-rater reliability coefficients using only Spearman's rho, which does not take into account the proportion of agreement due to chance.

The data support the internal consistency of the Shape Concern subscale and provide preliminary support for the internal consistency of the other three subscales. The highest internal consistencies were found in the samples of women with full- and sub-threshold AN and BN, whereas the lowest coefficients were found in the community-based or BED samples.

Finally, there is evidence that participants were able to accurately recall their symptom presentation as far back as 2.5 years. However, these data only examined whether participants accurately recalled the symptoms they reported at the prior interview and did not indicate whether participants accurately recalled the frequency of symptoms actually experienced.

Reliability of EDE-Q Scores

The data provide support for the test–retest reliability of the EDE-Q subscale scores and of scores on the following behavior frequency items: OBEs, self-induced vomiting, and laxative misuse. However, there is preliminary support for the test–retest reliability of scores on the items that assess the frequency of SBEs, OOE, and diuretic use. These data provide support for the temporal stability of the subscale scores over 5 to 14 months, but scores on the behavioral frequency items do not demonstrate temporal stability. It is notable that three of the four studies that examined the test–retest reliability and temporal stability of EDE-Q scores were conducted using community-based samples. When researchers only included participants who reported eating disorder symptoms, the test–retest correlations for OBEs were lower than when the analysis included the entire sample. Thus, the inclusion of respondents who report no disordered eating behavior may artificially inflate the correlations

between time 1 and time 2. Finally, the EDE-Q subscales demonstrate good internal consistency in both community samples of adult women and adult women with BN.

Validity of the EDE for the Assessment of Eating Disorder Symptoms

Overall, research indicates that EDE scores are able to distinguish between cases and noncases of eating disorders. One limitation of these data worth is that, in at least one study,¹⁹ it was unclear whether the assessors were blind to the participants' diagnostic status. On the basis of their percent of Ideal Body Weight (IBW), the women with AN weighed much less than the women with BN or the control women (73% IBW, 103% IBW, 100% IBW, respectively); thus, the assessors would likely be aware of the diagnostic status of the participants with AN. Thus, assessor knowledge of diagnostic status may limit the interpretability of these results. Despite this potential limitation, there appears to be support for the criterion-oriented validity of the EDE when it is used to differentiate between cases and noncases of eating disorders, even when the control participants report high restraint.

Research has demonstrated that scores on the subscales of the EDE correlate with measures of similar constructs. However, many of these correlations were only small to moderate and no studies have correlated EDE subscale scores with measures of similar constructs in AN, BED, male, or adolescent samples. Research has also found that frequencies of binge eating and compensatory behaviors reported on the EDE correlate with those reported on DFRs. However, only three studies have compared the EDE to DFRs and only one conducted a comparison of means analysis.

Three studies have examined the factor structure of the EDE, none of which replicated the EDE's four factor model. Although the results from the three studies were inconsistent, one consistent finding was that all three studies failed to discriminate between a Shape Concern factor and a Weight Concern factor. Because there was little overlap in the types of samples used, additional data are needed to determine whether different factor structures exist among participants with different symptom presentations.

Validity of the EDE-Q for the Assessment of Eating Disorder Symptoms

Overall, there are data to support the ability of scores on the EDE-Q to differentiate between cases and noncases of eating disorders. However, only one study used a structured interview to diagnose eating-disorder cases. One study found low concordance between the EDE-Q and another self-report measure in identifying regular binge eaters but both assessments were self-report questionnaires and it was unclear whether the discrepancy between the measures was a limitation of the EDE-Q, the other measure, or both. Additionally, all four studies compared the control samples to relatively small samples of eating disorder cases.

Research has demonstrated that scores on the Restraint subscale of the EDE-Q converge with scores on another measure of restraint compared to measures of bulimic symptoms. Additionally, the frequency of OBEs reported on the EDE-Q converge with the frequency of OBEs reported on DFRs in samples of adults with BED. However, there is only preliminary evidence that scores on the EDE-Q items measuring the frequency of SBEs and OOE converge with the frequency of these behaviors reported on DFRs. There are no published studies examining whether scores on the Eating Concern, Shape Concern, or Weight Concern sub-scales of the EDE-Q converge with measures of similar constructs. Likewise, no studies to date examine whether scores on the EDE-Q items that assess the frequencies of compensatory behaviors converge with the frequencies of these behaviors reported on DFRs, or using samples of adolescents or participants with AN, BN, or non-BED cases of EDNOS.

Studies using factor analysis to examine the factor structure of the EDE-Q provide moderate support for the construct validity of the Eating Concern and Restraint subscales in adult women with full-and sub-threshold BN. There was also moderate support for the Restraint subscale in bariatric surgery candidates. Consistent with the EDE, most of the questions from the EDE-Q Shape Concern and Weight Concern subscales loaded onto a single factor, suggesting that separating shape and weight may not be a meaningful distinction for many people. Finally, the data suggest that the importance of shape and weight represent a distinct construct and are not necessarily related to body dissatisfaction, discomfort with body exposure, or desire to change one's body shape and weight.

Conclusion

Current research provides support for the reliability and validity of scores on the EDE and EDE-Q for assessing eating disorder symptoms. However, there are notable limitations to this body of literature that need to be addressed in future research. First, there is an almost complete lack of research on the psychometric properties of the EDE and EDE-Q in males and adolescents. Second, there is a dearth of research on the psychometric properties of some portions of the EDE and EDE-Q (e.g., laxative and diuretic misuse, excessive exercise). It is also important to note that if these measures have demonstrated validity for a specific use (e.g., assessing Restraint) or in a specific sample (e.g., women with BN), this does not constitute evidence for the validity of the instruments for a different use (e.g., differentiating between eating disorder cases and noncases) or in another sample (e.g., women with BED). Third, the samples used appear to be primarily samples of convenience. For example, many studies appeared to use data collected from participants enrolled in treatment studies rather than collecting data with the specific purpose of examining the psychometric properties of the EDE or EDE-Q. Although research of this nature can provide important information regarding the psychometrics of these instruments, the generalizability of the findings is limited.

Research is also needed to further examine the validity of the EDE and EDE-Q with regard to the assessment of binge eating and compensatory behaviors. Given that specific frequencies of binge eating and compensatory behaviors are required to diagnose BN and BED, it is especially important to examine the sensitivity and specificity of the EDE and EDE-Q compared with DFRs. Additionally, as the diagnostic criteria for eating disorders evolve, the content validity of the EDE and EDE-Q should be evaluated. Finally, more rigorous research methods should be used to examine the psychometric properties of the EDE and EDE-Q, such as examining inter-rater reliability using either kappa coefficients or a T statistic⁴⁵ and examining construct validity using Multitrait-Multimethod matrices.⁴⁶

Acknowledgments

Supported by grants T32 MH082761-01 from NIMH and by P30DK 50456 from NIDDK.

References

1. Fairburn, CG.; Cooper, Z.; O'Connor, M. Eating disorder examination. In: Fairburn, CG., editor. *Cognitive Behavior Therapy and Eating Disorders*. 16. New York: Guilford Press; 2008. p. 265-308.
2. Fairburn, CG.; Cooper, Z. The eating disorder examination. In: Fairburn, CG.; Wilson, GT., editors. *Binge Eating: Nature, Assessment, and Treatment*. 12. New York: Guilford Press; 1993. p. 317-360.
3. Fairburn, CG.; Beglin, S. Eating disorder examination questionnaire. In: Fairburn, CG., editor. *Cognitive Behavior Therapy and Eating Disorders*. New York: Guilford Press; 2008. p. 309-313.

4. Fairburn CG, Beglin SJ. Assessment of eating disorders: Interview or self-report questionnaire? *Int J Eat Disord.* 1994; 16:363–370. [PubMed: 7866415]
5. Wilson, GT. Assessment of binge eating. In: Fairburn, CG.; Wilson, GT., editors. *Binge Eating: Nature, Assessment, and Treatment.* New York: Guilford Press; 1993. p. 227-249.
6. Grilo CM, Masheb RM, Wilson GT. A comparison of different methods for assessing the features of eating disorders in patients with binge eating disorder. *J Consult Clin Psychol.* 2001; 69:317–322. [PubMed: 11393608]
7. Reas DL, Grilo CM, Masheb RM. Reliability of the eating disorder examination-questionnaire in patients with binge eating disorder. *Behav Res Ther.* 2006; 44:43–51. [PubMed: 16301013]
8. Elder KA, Grilo CM. The spanish language version of the eating disorder examination questionnaire: Comparison with the spanish language version of the eating disorder examination and test-retest reliability. *Behav Res Ther.* 2007; 45:1369–1377. [PubMed: 17014823]
9. Grilo, CM. Structured instruments. In: Mitchell, JE.; Peterson, CB., editors. *Assessment of Eating Disorders.* New York: Guilford Press; 2005. p. 79-97.
10. Bryant-Waugh RJ, Cooper PJ, Taylor CL, Lask BD. The use of the eating disorder examination with children: A pilot study. *Int J Eat Disord.* 1996; 19:391–397. [PubMed: 8859397]
11. Grilo CM, Masheb RM, Lozano-Blanco C, Barry DT. Reliability of the eating disorder examination in patients with binge eating disorder. *Int J Eat Disord.* 2004; 35:80–85. [PubMed: 14705160]
12. Rizvi SL, Peterson CB, Crow SJ, Agras WS. Test-retest reliability of the eating disorder examination. *Int J Eat Disord.* 2000; 28:311–316. [PubMed: 10942917]
13. Luce K, Crowther JH. The reliability of the eating disorder examination–self-report questionnaire version (EDE-Q). *Int J Eat Disord.* 1999; 25:349–351. [PubMed: 10192002]
14. Cooper Z, Fairburn CG. The eating disorder examination: A semi-structured interview for the assessment of the specific psychopathology of eating disorders. *Int J Eat Disord.* 1987; 6:1–8.
15. Rosen JC, Vara L, Wendt S, Leitenberg H. Validity studies of the eating disorder examination. *Int J Eat Disord.* 1990; 9:519–528.
16. Masheb RM, Grilo CM. Rapid response predicts treatment outcomes in binge eating disorder: Implications for stepped care. *J Consult Clin Psychol.* 2007; 75:639–644. [PubMed: 17663617]
17. Beumont PJV, Kopec-Schrader EM, Talbot P, Touyz SW. Measuring the specific psychopathology of eating disorder patients. *Aust N Z J Psychiatry.* 1993; 27:506–511. [PubMed: 8250797]
18. Byrne SM, Allen KL, Lampard AM, Dove ER, Fursland A. The factor structure of the eating disorder examination in clinical and community samples. *Int J Eat Disord.* 2010; 43:260–265. [PubMed: 19350647]
19. Cooper Z, Cooper PJ, Fairburn CG. The validity of the eating disorder examination and its subscales. *Br J Psychiatr.* 1989; 154:807–812.
20. Grilo CM, Crosby RD, Peterson CB, Masheb RM, White MA, Crow SJ, et al. Factor structure of the eating disorder examination interview in patients with binge-eating disorder. *Obesity.* 2009; 18:977–981. [PubMed: 19798064]
21. Mond JM, Hay PJ, Rodgers B, Owen C, Beumont PJV. Temporal stability of the eating disorder examination questionnaire. *Int J Eat Disord.* 2004a; 36:195–203. [PubMed: 15282689]
22. Bardone-Cone AM, Boyd CA. The accuracy of symptom recall in eating disorders. *Compr Psychiatr.* 2007a; 48:51–56.
23. Peterson CB, Crosby RD, Wonderlich SA, Joiner T, Crow SJ, Mitchell JE, et al. Psychometric properties of the eating disorder examination-questionnaire: Factor structure and internal consistency. *Int J Eat Disord.* 2007b; 40:386–389. [PubMed: 17304585]
24. Ravaldi C, Vannacci A, Truglia E, Zucchi T, Mannucci E, Rotella CM, et al. The eating disorder examination as a retrospective interview. *Eat Weight Disord.* 2004; 9:228–231. [PubMed: 15656019]
25. Bardone-Cone AM, Agras WS. Psychometric properties of eating disorder instruments in black and white young women: Internal consistency, temporal stability, and validity. *Psychol Assess.* 2007; 19:356–362. [PubMed: 17845127]

26. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull.* 1955; 52:281–302. [PubMed: 13245896]
27. Reynolds CR. Measurement and assessment: An editorial view. *Psychol Assess.* 2010; 22:1–4. [PubMed: 20230146]
28. Wilfley DE, Schwartz MB, Spurrell EB, Fairburn CG. Using the eating disorder examination to identify the specific psychopathology of binge eating disorder. *Int J Eat Disord.* 2000; 27:259–269. [PubMed: 10694711]
29. Wilson GT, Smith D. Assessment of bulimia nervosa: An evaluation of the eating disorders examination. *Int J Eat Disord.* 1989; 8:173–179.
30. Greeno CG, Marcus MD, Wing RR. Diagnosis of binge eating disorder: Discrepancies between a questionnaire and clinical interview. *Int J Eat Disord.* 1995; 17:153–160. [PubMed: 7757096]
31. Wade T, Tiggemann M, Martin N, Heath A. A comparison of the eating disorder examination and a general psychiatric schedule. *Aust N Z J Psychiatry.* 1997; 31:852–857. [PubMed: 9483258]
32. Loeb KL, Pike KM, Walsh BT, Wilson GT. Assessment of diagnostic features of bulimia nervosa: Interview versus self-report format. *Int J Eat Disord.* 1994; 16:75–81. [PubMed: 7920584]
33. Barnes RD, Masheb RM, White MA, Grilo CM. Comparison of methods for identifying and assessing obese patients with binge eating disorder in primary care settings. *Int J Eat Disord.* 2011; 44:157–163. [PubMed: 20143322]
34. Spitzer, RL.; Yanovski, SZ.; Marcus, MD. *Questionnaire on Eating and Weight Patterns-Revised.* McLean, VA: BRS Search Service; 1993.
35. Farchus Stein K, Corte CM. Ecologic momentary assessment of eating-disordered behaviors. *Int J Eat Disord.* 2003; 34:349–360. [PubMed: 12949927]
36. Mannucci E, Ricca V, Di Bernardo M, Moretti S, Cabras PL, Rotella CM. Psychometric properties of the EDE 12. 0D in obese adult patients without binge eating disorder. *Eat Weight Disord.* 1997; 2:144–149. [PubMed: 14655838]
37. Elder KA, Grilo CM, Masheb RM, Rothschild BS, Burke-Martin-dale CH, Brody ML. Comparison of two self-report instruments for assessing binge eating in bariatric surgery candidates. *Behav Res Ther.* 2006; 44:545–560. [PubMed: 15993381]
38. Engelsen BK, Laberg JC. A comparison of three questionnaires (EAT-12, EDI, and EDE-Q) for assessment of eating problems in healthy female adolescents. *Nordic J Psychiatry.* 2001; 55:129–135.
39. Mond JM, Hay PJ, Rodgers B, Owen C, Beumont PJV. Validity of the eating disorder examination questionnaire (EDE-Q) in screening for eating disorders in community samples. *Behav Res Ther.* 2004b; 42:551–567. [PubMed: 15033501]
40. Wilson GT, Nonas CA, Rosenblum GD. Assessment of binge eating in obese patients. *Int J Eat Disord.* 1993; 13:25–33. [PubMed: 8477274]
41. Garner DM, Olmstead MP, Polivy J. Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. *Int J Eat Disord.* 1983; 2:15–34.
42. Garner DM, Garfinkel PE. Body image in anorexia nervosa: Measurement, theory, and clinical implications. *Psychol Med.* 1979; 9:273–279. [PubMed: 472072]
43. Grilo CM, Masheb RM, Wilson GT. Different methods for assessing the features of eating disorders in patients with binge eating disorder: A replication. *Obes Res.* 2001; 9:418–422. [PubMed: 11445665]
44. Hrabosky JL, White MA, Masheb RM, Rothschild BS, Burke-Martindale CH, Grilo CM. Psychometric evaluation of the eating disorder examination-questionnaire for bariatric surgery candidates. *Obesity.* 2008; 16:763–769. [PubMed: 18379561]
45. Lawlis GF, Lu E. Judgment of counseling process: Reliability, agreement, and error. *Psychol Bull.* 1972; 78:17–20. [PubMed: 5041757]
46. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull.* 1959; 56:81–105. [PubMed: 13634291]

TABLE 1
 Test-retest Reliability of Scores on the Eating Disorder Examination and Eating Disorder Examination-Questionnaire

References	N	Eating Disorder Examination									
		Restraint	Eating Concern	Shape Concern	Weight Concern	Objective Bulimic Days	Objective Bulimic Episodes	Subjective Bulimic Days	Subjective Bulimic Episodes	Vomiting Days	Vomiting Episodes
11 ^a	18	.88 ^{**}	.51 [*]	.50 [*]	.52 [*]	.71 ^{**}	.70 ^{**}	.17	.17	–	–
12 ^b	20	.76 ^{**}	.74 ^{**}	.76 ^{**}	.71 ^{**}	.85 ^{**}	.85 ^{**}	.40	.34	.97 ^{**}	.97 ^{**}
		Eating Disorder Examination-Questionnaire									
	N	Restraint	Eating Concern	Shape Concern	Weight Concern	Objective Bulimic Days	Objective Bulimic Episodes	Subjective Bulimic Days	Subjective Bulimic Episodes	Vomiting Days	Vomiting Episodes
13 ^a	139	.81 ^{**}	.87 ^{**}	.94 ^{**}	.92 ^{**}	.68 ^{**}	–	–	.92 ^{**}	.65 ^{**}	.54 ^{**}
7 ^b	86	.77 ^{**}	.72 ^{**}	.66 ^{**}	.71 ^{**}	.84 ^{**}	.51 ^{**}	.39 ^{**}	–	–	–

^a Pearson product moment correlation.

^b Spearman's rho.

* $p < .05$.

** $p \leq .001$.

TABLE 2

Inter-rater reliability coefficients for the Eating Disorder Examination

References	N	Restraint	Eating Concern	Shape Concern	Weight Concern	Objective Bulimic Days	Objective Bulimic Episodes	Subjective Bulimic Days	Subjective Bulimic Episodes	Vomiting Days	Vomiting Episodes
11 ^a	18	.96*	.90*	.84*	.65*	.99*	.98*	.91*	.91*	—	—
12 ^b	20	.95*	.94*	.90*	.99*	.99*	.99*	.99*	.91*	1.0*	1.0*
15 ^c	106	.92	.98	.99	.95	—	—	—	—	—	—

^aPearson product moment correlation.

^bSpearman's rho.

^cSignificance levels were not provided.

* $p < .001$.

TABLE 3
Internal Consistency of scores on the Eating Disorder Examination and Eating Disorder Examination-Questionnaire

References	N	Eating Disorder Examination		
		Restraint	Eating Concern	Shape Concern
17	116	.78	.68	.70
18 ^a	158	.64	.68	.85
18 ^b	170	.58	.69	.79
18 ^c	329	.65	.44	.77
19	142	.75	.78	.79
20	688	.63	.60	.68
Eating Disorder Examination-Questionnaire				
N	Restraint	Eating Concern	Shape Concern	Weight Concern
22 ^d	70	.81	.86	.89
22 ^e	156	.84	.85	.91
13 ^f	203	.84	.78	.93
13 ^g	139	.85	.81	.92
21 ^a	208	–	.73	.87
23	203	.70	.73	.83

^aEating disordered.

^bObese.

^cCommunity.

^dBlack women.

^eWhite women.

^fTime 1.

^gTime 2.

TABLE 4

Temporal stability of the Eating Disorder Examination-Questionnaire

References	N	Restraint	Eating Concern	Shape Concern	Weight Concern	Objective Bulimic Episodes	Subjective Bulimic Episodes	Objective Overeating Episodes	Excessive Exercise
22 ^a	70	.57 **	.79 **	.82 **	.81 **	-	-	-	-
22 ^b	156	.71 **	.81 **	.80 **	.81 **	-	-	-	-
21 ^a	196	.57 **	.77 **	.75 **	.73 **	.44 *	.28 *	.31 *	.31 *

^aBlack women.

^bWhite women.

* $p < .01$.

** $p \leq .001$.

TABLE 5
Cohen's d effect sizes for differences in Eating Disorder Examination Scores between groups

	Cooper et al. (1989) ¹⁹		Wilson and Smith (1989) ²⁹		Willfley et al. (2000) ²⁸	
	AN vs. NW	BN vs. NW	AN vs. BN	BN vs. NW	BED vs. NW	BED vs. OW
Restraint	1.83	2.06	0.02	0.40	0.94	0.16
Eating Concern	1.65	2.31	0.18	3.96	1.06	0.74
Shape Concern	2.16	2.64	0.54	4.87	3.50	1.35
Weight Concern	1.64	2.34	0.50	6.68	3.86	1.69
Objective Bulimic Episodes	0.62	1.34	0.62	–	2.55	2.54
Self-Induced Vomiting	0.62	1.22	0.33	–	1.86	1.87

Notes: AN = anorexia nervosa; NW = normal weight control; BN = bulimia nervosa; BED = binge eating disorder; OW = overweight control.

TABLE 6
 Convergence of the Eating Disorder Examination Subscales and measures of similar constructs

	Loeb et al. (1994) ³² (N = 82)				Rosen et al. (1990) ¹⁵ (N = 106)			
	Restraint	Eating Concern	Shape Concern	Weight Concern	Restraint	Eating Concern	Shape Concern	Weight Concern
Caloric Size of Binge Episodes	-	-	-	-	-	.52***	-	-
Frequency of Binge Eating	-	-	-	-	-	.50***	-	-
Frequency of Regular Meals	-	-	-	-	-.37***	-	-	-
Frequency of Snack Foods	-	-	-	-	-.22**	-	-	-
Nonvomited Caloric Intake	-	-	-	-	-.39***	-	-	-
BSQ	-	-	.76*	.61*	-	-	.82***	.78***
EAT Bulimia and Food Preoccupation	-	.35**	-	-	-	-	-	-
EAT Dieting	.54*	.37*	.36*	.35**	-	-	-	-
EAT Oral Control	.22***	-	-	-	-	-	-	-
TFEQ Restraint	.48*	-	-	-	-	-	-	-

Notes: BSQ = Body Shape Questionnaire; EAT = Eating Attitudes Test; TFEQ = Three-Factor Eating Questionnaire.

* $p < .05$.

** $p \leq .01$.

*** $p \leq .001$.

TABLE 7
 Convergence of the Eating Disorder Examination and daily food records for binge eating and self-induced vomiting

References	N	Binge Eating			
		Correlation between EDE (7 Days) and DFR (7 Days)	Correlation between EDE (7 Days) and DFR (28 Days)	Correlation between EDE (28 Days) and DFR (7 Days)	Correlation between EDE (28 Days) and DFR (28 Days)
35	13	–	–	–	.60*
32 ^a	69	.88**	–	.90**	–
32 ^b	50–52 ^c	.87**	.80**	.91**	.93**
15	106	–	–	.56**	–
Self-Induced Vomiting					
References	N	Self-Induced Vomiting			
		Correlation between EDE (7 Days) and DFR (7 Days)	Correlation between EDE (7 Days) and DFR (28 Days)	Correlation between EDE (28 Days) and DFR (7 Days)	Correlation between EDE (28 Days) and DFR (28 Days)
35	13	–	–	–	.75**
32 ^a	59–69 ^c	.88**	–	.93**	–
32 ^b	50–52 ^c	.98**	.97**	.95**	.99**
15	106	–	–	.90**	–

Notes: EDE = Eating Disorder Examination; DFRs = Daily Food Records.

^a Pretreatment.

^b Post-treatment.

^c Ns vary due to missing data.

* $p < .01$.

** $p < .001$.