Published in final edited form as:

Psychol Assess. 2011 September ; 23(3): 714–724. doi:10.1037/a0023246.

Convergence of Scores on the Interview and Questionnaire Versions of the Eating Disorder Examination: A Meta-Analytic Review

Kelly C. Berg,

Department of Psychology, University of Minnesota, Department of Psychiatry, University of Minnesota

Carol B. Peterson,

Department of Psychiatry, University of Minnesota

Patricia Frazier, and

Department of Psychology, University of Minnesota

Scott J. Crow

Department of Psychiatry, University of Minnesota.

Abstract

Significant discrepancies have been found between interview- and questionnaire-based assessments of psychopathology; however, these studies have typically compared instruments with unmatched item content. The Eating Disorder Examination (EDE), a structured interview, and the questionnaire version of the EDE (EDE-Q) are considered the preeminent assessments of eating disorder symptoms and provide a unique opportunity to examine the concordance of interview- and questionnaire-based instruments with matched item content. The convergence of EDE and EDE-Q scores has been examined previously; however, past studies have been limited by small sample sizes and have not compared the convergence of scores across diagnostic groups. A meta-analysis of 16 studies was conducted to compare the convergence of EDE and EDE-Q scores across studies and diagnostic groups. With regard to the EDE and EDE-Q subscale scores, the overall correlation coefficient effect sizes ranged from .64 to .75. The overall Cohen's *d* effect sizes ranged from .31 to .59 with participants consistently scoring higher on the questionnaire. With regard to the items measuring behavior frequency, the overall correlation coefficient effect sizes ranged from .49 to .64 for binge eating and .84 to .89 for compensatory behaviors. The overall Cohen's *d* effect sizes ranged from -.14 to -.23, with participants reporting more binge eating on the interview in 70% of the studies. These results suggest that the interview and questionnaire assess similar constructs, but that the two instruments should not be used interchangeably. Additional research is needed to examine the inconsistencies between binge frequency scores on the two instruments.

Correspondence concerning this article should be addressed to Kelly C. Berg, Department of Psychiatry, University of Minnesota, 606 – 24th Avenue South, Suite 602, Minneapolis, MN 55454. bergx143@umn.edu.

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/pubs/journals/pas

This manuscript is based, in part, on analyses conducted for an unpublished doctoral dissertation.

Keywords

Eating Disorder Examination; Eating Disorder Examination-Questionnaire; Convergent validity; Meta-analysis

Interviews and questionnaires are two common methods of assessing symptoms of eating disorders, and, as in other areas of mental health, they are often used interchangeably depending on available resources. Additionally, results from interview- and questionnaire-based instruments are regularly compared across studies and few researchers comment on the modality of the assessments as the two types of instruments theoretically assess the same constructs and should be highly concordant. However, research on the assessment of depression (Reeves, Large, & Honeyman, 1985), personality disorders (Angus & Marziali, 1988; Zimmerman & Coryell, 1990), panic disorder (Woodruff-Borden, Jeffery, Bourland, Brothers, & Albano, 2000), and chronic pain (Campbell, Quilty, & Dieppe, 2003) has demonstrated poor agreement between interview- and questionnaire-based assessments. The low levels of concordance between the two types of assessments call into question whether interview- and questionnaire-based instruments do measure similar constructs when used to measure symptoms of psychopathology. However, most of the studies cited above compared interviews and questionnaires that were not necessarily matched on item content. The purpose of this study was to examine the concordance of scores on interview- and questionnaire-based instruments with matched item content in the domain of eating disorder assessment.

Eating Disorder Examination

The Eating Disorder Examination (EDE; Fairburn & Cooper, 1993; Fairburn, Cooper, & O'Connor, 2008) is a semi-structured interview that assesses the cognitive and behavioral symptoms of eating disorders and has been referred to as the most comprehensive (Wilson, 1993) and most thoroughly evaluated (Grilo, 2005) eating disorder assessment. The EDE includes four subscales that purport to measure cognitive features of eating disorders: Restraint, Eating Concern, Shape Concern, and Weight Concern. The EDE also asks respondents to estimate the frequency with which behaviors associated with eating disorders have occurred during the past three (and six) months. These include Objective Bulimic Episodes (OBEs), Subjective Bulimic Episodes (SBEs), self-induced vomiting, laxative misuse, diuretic misuse, and excessive exercise. The breadth of symptoms assessed by the EDE allows the instrument to be used with individuals presenting with any of the four eating disorders currently included in the Diagnostic and Statistical Manual for Mental Disorders (DSM), 4th edition, Text Revision (American Psychiatric Association [APA], 1994): Anorexia Nervosa (AN), Bulimia Nervosa (BN), Binge Eating Disorder (BED)¹, and Eating Disorder Not Otherwise Specified (EDNOS). Additionally, data from the EDE can be analyzed in multiple ways. For example, subscale scores and behavior frequency scores can be used as dimensional variables (e.g., Fairburn, Cooper, Doll, Norman, & O'Connor, 2000; Le Grange, Crosby, Rathouz, & Leventhal, 2007) whereas individuals' diagnostic status, as determined by algorithms provided by Fairburn and Cooper (1993), can provide categorical information (e.g., Agras, Walsh, Fairburn, Wilson, & Kraemer, 2000; Peterson, Mitchell, Crow, Crosby, & Wonderlich, 2009). Its status as the gold standard of eating disorder assessment has also given the EDE the weighty responsibility of serving to validate other assessments (Grilo, Masheb, & Wilson, 2001a; Reas, Grilo, & Masheb, 2006).

¹Binge Eating Disorder is included in Appendix B as a criterion set suggested for further study (APA, 1994).

Unfortunately, the EDE is lengthy to administer and requires significant amounts of assessor training. A questionnaire version of the EDE (EDE-Q) was developed to address these limitations by attempting to measure the same constructs as the EDE in a self-report questionnaire format (Fairburn & Beglin, 1994, 2008). The EDE-Q includes the same items used to generate the four subscales and to determine the frequency of binge eating and compensatory behaviors. Additionally, the EDE-Q items are worded almost identically to those in the EDE. The primary difference between the EDE and EDE-Q is that the questionnaire is based only on a 28-day time period and the interview allows a trained assessor to clarify concepts and ask additional questions.

The psychometric properties of the EDE and EDE-Q have been examined in depth elsewhere (Berg, Peterson, Frazier, & Crow, 2010). To summarize, with the exception of SBE frequency, scores on both instruments have demonstrated acceptable test-retest reliability (range of $r_s = .51$ to $.97$; e.g., Grilo, Masheb, Lozano-Blanco, & Barry, 2004; Reas et al., 2006) and internal consistency (range of mean $\alpha_s = .65$ to $.89$; e.g., Grilo et al., 2009; Peterson et al., 2007). Research also supports the inter-rater reliability of EDE scores (Grilo et al., 2004). Additionally, both the EDE and EDE-Q have demonstrated the ability to distinguish between eating disorder and non-eating disorder cases (e.g., Cooper, Cooper, & Fairburn, 1989; Mond, Hay, Rodgers, Owen, & Beumont, 2004) and the subscale scores of these instruments correlate with scores on measures of similar constructs (e.g., Grilo, Masheb, & Wilson, 2001a; Loeb, Pike, Walsh, & Wilson, 1994).

Convergence of Scores on the EDE and EDE-Q

As stated earlier, the EDE and EDE-Q use the same items to generate the Restraint, Eating Concern, Shape Concern, and Weight Concern subscales as well as to determine the frequency of binge eating and compensatory behaviors. Although researchers have compared scores on the two instruments (e.g., Binford, Le Grange, & Jellar, 2005; Fairburn & Beglin, 1994; Grilo, Masheb, & Wilson, 2001a), there are important limitations to this body of literature. First, most of these studies have used correlations and significance testing to compare the two instruments. Unfortunately, correlations can only indicate whether there is a relationship between scores on two measures. These relationships may exist in the presence *or* absence of significant differences between mean scores on the two measures. Significance testing is also limited because it is based on both the size of the effect and the size of the sample. Without a better understanding of the strength of the relationships and differences between scores on the interview and questionnaire, it is impossible to know whether the two instruments yield similar conclusions regarding symptom presentation. Second, most of the studies used small samples which limit the power as well as the generalizability of the results. Lastly, most of the published studies have compared the interview and questionnaire versions of the EDE in specific diagnostic groups (e.g., AN, BN), but none have compared the convergence of the two instruments across diagnostic groups.

The purpose of this study was to examine the concordance of interview- and questionnaire based instruments with matched item content by analyzing the convergence of scores on the EDE and EDE-Q using a meta-analytic strategy. Cohen's d (Cohen, 1988) effect sizes were calculated to compare mean scores across the EDE and EDE-Q and correlation coefficients were used to determine the overall strength of the relationship between EDE and EDE-Q scores across studies. Finally, a homogeneity analysis was conducted on both types of effect sizes to compare effect sizes across studies.

Method

Literature Search Procedure and Study Eligibility

Between July, 2009 and July, 2010, multiple literature searches were conducted for published studies that assessed the convergence of EDE and EDE-Q scores using three major computer databases (i.e., MEDLINE, PsycINFO, PubMed) and reviewing reference lists from published journal articles and books. Search terms used included “Eating Disorder Examination” and “Eating Disorder Examination-Questionnaire.” Studies were included that administered the EDE and EDE-Q at the same time point and assessed the convergence of scores on the EDE and EDE-Q using correlation coefficients or by comparing mean scores. The literature search was inclusive of studies that examined the convergence of scores for any of the four subscales (i.e., Restraint, Eating Concern, Shape Concern, or Weight Concern), binge eating (i.e., frequency of OBEs or SBEs), or compensatory behaviors (i.e., frequency of self-induced vomiting, laxative misuse, diuretic misuse, or excessive exercise). Studies were excluded if the research was published in a language other than English, if the study used a nontraditional administration of the instrumentsⁱⁱ, or if translated (e.g., Becker et al., 2010; Grilo, Lozano, & Elder, 2005), or child (e.g., Bryant-Waugh, Cooper, Taylor, & Lask, 1996) versions of the EDE or EDE-Q were used.

If an eligible study did not include means and standard deviations for the EDE and EDE-Q or if correlations between the EDE and EDE-Q were not reported, the first author attempted to contact the primary author to obtain these statistics. Of the three authors contacted, all three responded and provided data for three of the four studies that had missing data. An eligible study was excluded from the meta-analysis only if the statistics necessary to conduct the meta-analysis (e.g., means, standard deviations, correlation coefficients) were not reported or could not be obtained. If an eligible study was excluded from the meta-analysis using Cohen's d , it could have been used for the meta-analysis using correlation coefficients and vice versa.

Statistical Methods

The meta-analysis included three parts: 1) a meta-analysis examining the relationship between scores on the EDE and EDE-Q using correlation coefficient effect sizes, 2) a meta-analysis examining differences between mean scores on the EDE and EDE-Q using Cohen's d effect sizes, and 3) a homogeneity analysis to examine differences in effect sizes across diagnostic groups. Separate effect sizes were calculated for scores on each variable in each applicable subsample. For example, if a study reported correlations between EDE and EDE-Q scores for the frequency of OBEs and self-induced vomiting in both a BN sample and a community sample, separate effect sizes were calculated for the frequency of OBEs reported in the BN sample and in the community sample and for the frequency of self-induced vomiting in the BN sample and in the community sample.

Calculation of correlation coefficient effect sizes—Correlation coefficient effect sizes were calculated using the procedure and formulas detailed by Lipsey and Wilson (2001). After all applicable correlation coefficients were identified, each correlation coefficient was standardized using Fisher's Z_r transformation and then weighted by its inverse variance weight to adjust for sample size. Mean weighted correlation coefficient effect sizes ($MWES_{Z_r}$) were calculated by dividing the sum of the weighted effect sizes by the sum of the inverse variance weights. The $MWES_{Z_r}$ was calculated for each variable in

ⁱⁱOne study was found that examined the convergence of the EDE and EDE-Q when the EDE was conducted by phone and the EDE-Q was administered online (Pretorius, Waller, Gowers, & Schmidt, 2009). As it is yet unclear how nontraditional administrations of the EDE and EDE-Q affect response patterns, this study was not included in the meta-analysis.

each applicable subgroup (e.g., the subgroup of studies that reported correlations between EDE and EDE-Q scores on the Restraint subscale in AN samples) as well as the total $MWES_{Z_r}$ for each variable across all studies (e.g., all studies that reported correlations between EDE and EDE-Q scores on the Restraint subscale). Each $MWES_{Z_r}$ was then reverse transformed to $MWES_r$. Finally, confidence intervals (CI) for each $MWES_r$ were calculated, first using $MWES_{Z_r}$ and then reverse transformed from Z_r to r .

Calculation of Cohen's d —The calculation for Cohen's d also followed the steps outlined by Lipsey and Wilson (2001). Cohen's d was calculated by subtracting the mean EDE from the mean EDE-Q score and then dividing the result by the pooled standard deviation of the EDE and EDE-Q scores. As a result, positive Cohen's d coefficients indicate higher scores on the self-report questionnaire (EDE-Q) whereas negative Cohen's d coefficients indicate higher scores on the interview (EDE). Consistent with procedures used for the correlation coefficient effect sizes, each Cohen's d was weighted to adjust for sample size by multiplying each Cohen's d by its respective inverse variance weight. Mean weighted effect sizes for Cohen's d ($MWES_d$) were calculated by dividing the sum of the weighted effect sizes by the sum of the inverse variance weights. The $MWES_d$ was calculated for each variable in each applicable subgroup as well as across all studies, as described above. The confidence intervals (CI) for each $MWES_d$ were also calculated. In this study, small Cohen's d coefficients represent small differences between the EDE and EDE-Q and thus higher convergence between the two instruments.

Homogeneity analysis—Meta-analytic techniques assume that all effect sizes used in the meta-analysis are estimated from the same population. A homogeneity analysis tests whether this assumption holds true. If the homogeneity assumption is rejected, the effect sizes are estimates of at least two populations. This type of analysis is of particular importance to this study as this meta-analysis purposefully included studies that sampled different populations (e.g., AN samples, community samples). Thus, the homogeneity assumption was tested for both the meta-analyses using correlation coefficients and Cohen's d to determine whether the means and the relationship between scores on the two instruments was similar across diagnostic groups.

The homogeneity assumption was tested using the Q -test which is distributed as a chi-square statistic. To account for both sampling error and variability in population effects, a random effects model was used (Lipsey & Wilson, 2001). Given that the Q -test may have limited power to detect heterogeneity when the number of studies included in the meta-analysis is small, I^2 was also calculated because it is not sample-size dependent (Higgins & Thompson, 2002). I^2 indicates the proportion of variability in effect sizes due to heterogeneity.

Results

Study Characteristics

Sixteen studies (Barnes, Masheb, White, & Grilo, in press; Binford et al., 2005; Black & Wilson, 1996; Carter, Aime, & Mills, 2001; de Zwaan et al., 2004; Fairburn & Beglin, 1994; Goldfein, Devlin, & Kamenetz, 2005; Grilo, Masheb, & Wilson, 2001a; Grilo, Masheb, & Wilson, 2001b; Kalarchian, Wilson, Brodin, & Bradley, 2000; Mond et al., 2004; Passi, Bryson, & Lock, 2003; Sysko, Walsh, & Fairburn, 2005; Sysko, Walsh, Schebendach, & Wilson, 2005; Wilfley, Schwartz, Spurrell, & Fairburn, 1997; Wolk, Loeb, & Walsh, 2005) were identified that met the inclusion and exclusion criteria described above. The diagnostic groups represented by the sixteen studies were as follows: AN (4), BN (3), combined AN and BN sample (1), EDNOS (1), Binge Eating Disorder (5), community sample (2), bariatric surgery patients (2), substance abusers (1). Three of the studies included adolescent samples

($n = 111$, mean age = 15.6) and 14 included adult samples ($n = 1,173$, mean age = 33.9). Of the 1,284 participants included in these studies, approximately 94% were female, and 62% were either treatment-seeking or enrolled in treatment during the study. Table 1 summarizes the characteristics of each study.

Of the 16 studies included in the meta-analysis, 12 administered the EDE-Q prior to the EDE, one administered the EDE prior to the EDE-Q (Carter et al., 2001), two did not control the order of administration (de Zwaan et al., 2004; Sysko, Walsh, & Fairburn, 2005), and one did not report on the order of administration (Sysko, Walsh, Schebendach, & Wilson, 2005). The majority of the studies chose to administer the questionnaire prior to the interview rather than counterbalancing them because definitions of key variables are purposefully elaborated on during the interview. Two studies have found that administering the EDE-Q after the EDE results in higher correspondence between the two instruments than administering the interview after the questionnaire (Carter et al., 2001; Passi et al., 2003). Because participants who completed the EDE-Q after the EDE would have received more comprehensive explanations of “binge eating” and “loss of control,” these data support the hypothesis that giving respondents additional information regarding the definitions of key terminology may enhance the correspondence between the EDE and EDE-Q. Thus, it is likely that administering the interviewer prior to the questionnaire would bias the responses to the questionnaire whereas it is unlikely that administering the questionnaire first would bias scores on the interview (Fairburn & Beglin, 1993).

Subscale Scores

Meta-analytic results for the behavior frequency items are reported in Tables 2 and 3. The pattern of findings was consistent across the four subscales (Restraint, Eating Concern, Shape Concern, Weight Concern) and showed good convergence of scores on the subscales across the EDE and EDE-Q (r 's = .68 to .76) but small to medium differences in means (Cohen's d 's = .31 to .62). Participants scored higher on the EDE-Q than the EDE on all four subscales (see Tables 2 and 3).

With regard to the meta-analysis using correlation coefficients, the Q -test did not reach statistical significance for any of the subscales suggesting that the variability in effect sizes did not exceed what would be expected given errors in sampling. However, for the Eating Concern, Shape Concern, and Weight Concern subscales, I^2 ranged from 19% to 33%. In contrast, the I^2 for the Restraint subscale was 0%, indicating that all variability in effect sizes was due to sampling error. Similarly, with regard to the meta-analysis using Cohen's d , the Q -test did not reach significance for any of the subscales; however, I^2 ranged from 11% to 30% across the four subscales, indicating that a small amount of the variability in effect sizes was due to heterogeneity.

Behavior Frequency Items

Meta-analytic results for the behavior frequency items are reported in Tables 4 and 5. The results demonstrated lower convergence of EDE and EDE-Q scores for the items that measure the frequency of OBEs and SBE's (r 's = .55 and .37, respectively) as well as small differences in mean scores (d 's = -.16 and -.22, respectively). On these scales, participants scored higher on the EDE than on the EDE-Q.

There was strong convergence on EDE and EDE-Q scores for items that measure the frequency of self-induced vomiting ($r = .90$) and laxative misuse ($r = .92$). However, because only two studies (Carter et al., 2001; Wolk et al., 2005) reported the means and standard deviations for self-induced vomiting and laxative misuse, a meta-analysis using Cohen's d was considered inappropriate. The d 's in these studies ranged from -.35 to .09.

Only two studies reported data on compensatory behaviors other than self-induced vomiting and laxative misuse (Carter et al., 2001; Wolk et al., 2005). Of these, one compared the frequency of diuretic misuse reported on the EDE and EDE-Q ($d = .01$; Carter et al., 2001) and the other compared the frequency of excessive exercise reported on the two instruments ($d = .15$; Wolk et al., 2005). Again, the dearth of research precludes the use of meta-analysis to compare scores on the EDE and EDE-Q for these variables.

With regard to the meta-analysis using correlation coefficients, the Q -test did not reach statistical significance for any of the behavior frequency items. Additionally, the I^2 values of 0% for OBEs, SBEs, and laxative misuse suggest that variability in effect sizes were due to sampling error. In contrast, the I^2 of 20% for self-induced vomiting indicates that some of the variability in effect sizes was due to heterogeneity. Regarding the meta-analysis using Cohen's d , the Q -test did not reach statistical significance for either OBEs or SBEs. The I^2 value for OBEs (0%) also suggests that the variability in effect sizes was due to sampling error whereas the I^2 value for SBEs (19%) indicates that a small amount of the variability in effect sizes was due to between-group variability.

Discussion

The purpose of this study was to examine the convergence of scores on interview- and questionnaire-based instruments using the domain of eating disorder assessment as an example. The interview and questionnaire versions of the EDE are nearly identical with regard to item content, item wording, and scoring, which provides a unique opportunity to examine differences in response patterns between interview- and questionnaire-based assessments of psychopathology. This is the first study to examine the convergence of scores on the interview and questionnaire versions of the EDE using meta-analysis. Given the small sample sizes used in most previous research in this area, meta-analysis is essential to understanding the generalizability of the results. Additionally, both meta-analysis using correlation coefficients and Cohen's d were used which allows for interpretation of both the relationship between scores on the two instruments and the size of the difference between mean scores on the two instruments. Finally, a homogeneity analysis was used to examine whether the degree of concordance between scores on the EDE and EDE-Q is consistent across different diagnostic groups.

The results from the meta-analysis provide support for the convergence of scores on the interview and questionnaire versions of the EDE. However, the degree of concordance varies, particularly for the Eating Concern, Shape Concern, and Weight Concern subscales and for the self-induced vomiting item. The results from the meta-analysis using Cohen's d effect sizes show that there are small to medium effects for the differences between mean scores on the EDE and EDE-Q for the Restraint subscale, Weight Concern subscale, and the frequency of OBEs and SBEs. Additionally, there were medium to large effect sizes for the differences between the EDE and EDE-Q for the Eating Concern and Shape Concern subscales. The effect sizes vary between studies for all four subscales and SBE frequency. In contrast, variability in effect sizes for OBE frequency was due entirely to sampling error. These findings have important clinical implications for the assessment of eating disorder symptoms, as discussed below.

Clinical Implications

Cognitive symptoms—The similarity between the EDE and EDE-Q with regard to item content provides a unique opportunity to compare response patterns between self-report and interview-based measures of psychopathology. With regard to the four subscales of the EDE and EDE-Q that measure cognitive symptoms of eating disorders, as mentioned, the results of the meta-analysis indicate that participants who report high levels of symptoms on one of

the two instruments also report high levels of symptoms on the other. However, the results also demonstrate that participants consistently report higher levels of symptoms on the questionnaire version of the EDE than during the interview. These results imply that participants either over-report cognitive symptoms on the questionnaire or under-report these symptoms during the interview. It has been suggested that respondents may under-report their symptoms during interviews because of feelings of shame elicited by the loss of anonymity during face-to-face interviews (e.g., Grilo, 2005). This hypothesis has been supported by the findings that participants were more likely to endorse eating disorder symptoms under conditions that were characterized by higher perceived anonymity (e.g., using an unmatched count response format; Lavender & Anderson, 2009) and that questionnaire and interview scores were more similar when interviews were conducted via telephone rather than in person (Keel, Crow, Davis, & Mitchell, 2002).

Alternatively, respondents may under-report their symptoms during interviews not because of shame, but because their symptoms are not perceived as problematic and they do not want treatment. There is empirical support for this theory as one study found that women who endorsed purging behavior on the questionnaire version of the EDE and subsequently denied this behavior during the interview were significantly less functionally impaired and distressed than women who endorsed purging behavior on both instruments (Mond, Hay, Rodger, & Owen, 2007).

It should be noted that the studies described above used primarily community samples or non-treatment seeking samples. In contrast, most of the participants included in this meta-analysis were either seeking treatment or already enrolled in treatment. This consideration is important because research from the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kraemmer, 1989), a self-report questionnaire, has demonstrated that demoralization or distress can elevate scores on the Clinical Scales and the Infrequency Scale (F) over and above those typically observed in psychiatric samples (Arbisi & Ben-Porath, 1995; Sellbom, Ben-Porath, McNulty, Arbisi, & Graham, 2006). Additionally, research on the concordance between interview- and questionnaire-based assessments of personality disorder symptoms found that high levels of depression were positively correlated with larger discrepancies between the two assessments (Zimmerman & Coryell, 1990). Given that patients typically seek treatment when in distress, patients seeking treatment may over-report their symptoms on self-report questionnaires due to high levels of distress. Although one might argue that distress would also inflate participants' scores on interviews, semi-structured interviews such as the EDE provide anchors that assessors can use to make ratings, thereby decreasing the bias caused by participant distress (Wilson, 1993).

Although subscale scores were higher on the EDE-Q than the EDE, it is notable that the difference between the two measures was greater for the Eating Concern and Shape Concern subscales than for the Restraint and Weight Concern subscales. This finding is particularly interesting given that the distinction between shape and weight on the EDE may not be empirically supported (e.g., Byrne, Allen, Lampard, Dove, & Fursland, 2010; Grilo et al., 2009). It is possible that the variable responsible for higher scores on the EDE-Q (e.g., shame, distress, etc.) has a greater impact on the Eating Concern and Shape Concern subscales than on the Restraint or Weight Concern scales; however, further research is needed to examine this question.

Behavioral symptoms—The correlations between scores on the EDE and EDE-Q for the assessment of compensatory behavior frequency were very high (r 's = .90 to .92) whereas the correlation between scores on the two instruments for the assessment of binge eating frequency were much lower (r 's = .37 to .55). Thus, there is a stronger relationship between

scores on the two instruments with regard to the assessment of compensatory behaviors than for the assessment of binge eating. Although there were only small differences between mean scores on the EDE and EDE-Q with regard to the assessment of OBEs and SBEs, these data do not necessarily support the convergence of scores on the EDE and EDE-Q for the assessment of binge eating frequency. As stated previously, the correlations between scores on the two instruments were lower for the assessment of binge eating frequency than for the assessment of compensatory behavior frequency or cognitive symptoms. Additionally, the range of Cohen's *d* effect sizes was large, ranging from -.58 to .26 for OBEs and -.57 to .17 for SBEs, indicating that participants did not consistently score higher on one instrument than the other. Thus, the small observed Cohen's *d* may be the result of an averaging out of positive and negative effect sizes. Overall, these data indicate that there are inconsistencies between the frequency of binge eating reported on the interview and questionnaire versions of the EDE that do not exist for the assessment of cognitive symptoms or the frequency of compensatory behaviors. This is problematic given that diagnoses of BN and BED require the presence of specific frequencies of binge eating.

It has been suggested that the inconsistencies between interview and questionnaire-based instruments used to assess binge eating frequency may be due to the vague, ambiguous definition of binge eating and that giving participants more information regarding the definitions of binge eating may increase the accuracy with which participants report these behaviors on self-report questionnaires such as the EDE-Q (Wilfley et al., 1997). Based on this hypothesis, the Eating Disorder Examination–Questionnaire was modified to include definitions of a “large amount of food” and “loss of control.” The limited amount of research on the Eating Disorder Examination–Questionnaire with Instructions (EDE-Q-I; Goldfein et al., 2005) has found that binge eating frequency scores on the EDE-Q-I correlate more strongly with the frequency of binge eating reported on the EDE than do scores on the original EDE-Q (Celio, Wilfley, Crow, Mitchell, & Walsh, 2004; Goldfein et al., 2005).

Conclusions and Limitations

In sum, the results from this study generally support the convergence of scores on the interview and questionnaire versions of the EDE and suggest that the two instruments measure similar constructs. This support is strongest for the Restraint, Eating Concern, Shape Concern, and Weight Concern subscales as well as for the assessment of self-induced vomiting and laxative misuse. These data provide more limited support for the convergence of scores on interview- and questionnaire-based assessments of binge eating frequency. Additionally, the results suggest that there may be differences in the amount of concordance between interview- and questionnaire-based instruments depending on the type of sample being used and the symptom being assessed.

Clinically, these results suggest that both the interview and questionnaire versions of the EDE can be used to assess the cognitive and behavioral symptoms of eating disorders. However, given the inconsistencies between the frequency of binge eating reported on the EDE and EDE-Q, these data do not support using the two instruments interchangeably as differences in symptom levels between the two instruments may be erroneously attributed to factors such as time or treatment condition when they should in fact be attributed to differences in the modality of the assessment.

This study has several strengths. The interview and questionnaire versions of the EDE are nearly identical with regard to item content, item wording, and scoring, which provides a unique opportunity to examine differences in response patterns between interview- and questionnaire-based assessments of psychopathology. Second, this is the first study to examine the convergence of scores on the interview and questionnaire versions of the EDE using meta-analysis. Given the small sample sizes used in most previous research in this

area, meta-analysis is essential to understanding the generalizability of the results. Additionally, both meta-analysis using correlation coefficients and Cohen's *d* were used which allows for interpretation of both the relationship between scores on the two instruments and the size of the difference between mean scores on the two instruments. Third, a homogeneity analysis was used to examine whether the degree of concordance between scores on the EDE and EDE-Q is consistent across different diagnostic groups.

This study also had several limitations. First, there are important limitations to the studies that were included in the meta-analysis. Overall, there is a dearth of research on the relationship between the EDE and EDE-Q for participants with BN and EDNOS, males, adolescents, and for the assessment of compensatory behavior frequency. It is also notable that 75% of the studies purposefully chose not to counterbalance the administration of the EDE and EDE-Q and administered the questionnaire first. As mentioned earlier, the rationale for this decision is that, because the EDE provides detailed definitions of key variables, the interview is more likely to bias responses on the questionnaire than vice versa (Fairburn & Beglin, 1993). We were unable to statistically examine or control for the order of administration in the current meta-analysis because of lack of power. The one study that administered the interview first was the only study to consistently find higher subscale scores on the EDE than the EDE-Q (Carter et al., 2001). However, given that the sample size was one, it is unclear whether this finding was due to ordering effects or simply chance.

Second, although the EDE and EDE-Q are well regarded measures of eating disorder pathology, they are not without flaws. For example, there is a lack of data supporting the original factor structure of the EDE and EDE-Q (Berg et al., manuscript accepted for publication). The lower test-retest reliability and internal consistency of the scores on some of the subscales also place constraints on the extent to which the EDE and EDE-Q can correlate.

Finally, there are also limitations to the meta-analysis itself. First, this study does not provide information regarding the convergence of individual symptom profiles between the EDE and EDE-Q and future researchers may consider examining whether these measures arrive at similar diagnostic conclusions. Additionally, the results from the meta-analysis can only be used to describe the relationship between the two instruments. These data do not provide evidence as to the cause of the differences between the EDE and EDE-Q; thus, specific recommendations regarding which assessment tool is preferable and whether one instrument may be preferable for use in certain samples or under certain circumstances can only be made pending additional research. Future research should continue to explore whether self-report questionnaires over-estimate symptom levels or whether interview-based assessments under-estimate symptom levels. Researchers may find that both over-reporting and under-reporting occur, in which case it may be important to understand both the assessment conditions and participant variables that contribute to both.

Acknowledgments

This research was supported, in part, by grants from NIMH (T32 MH082761-01) and NIDDK (P30DK 50456).

References

References marked with an asterisk indicate studies included in the meta-analysis.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders, fourth edition, text revision (DSM-IV-TR). APA; Washington, D.C.: 1994.

- Agras WS, Walsh BT, Fairburn CG, Wilson GT, Kraemer HC. A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa. *Archives of General Psychiatry*. 2000; 57:459–466. [PubMed: 10807486]
- Angus LE, Marziali E. A comparison of three measures for the diagnosis of borderline personality disorder. *The American Journal of Psychiatry*. 1988; 145:1453–1454. [PubMed: 3189609]
- Arbisi PA, Ben-Porath YS. An MMPI-2 infrequent response scale for use with psychopathological populations: The infrequency-psychopathology scale, F(p). *Psychological Assessment*. 1995; 7:424–431.
- *. 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. *International Journal of Eating Disorders*. (in press).
- Becker AE, Thomas JJ, Bainivualiku A, Richards L, Navara K, Roberts AL, Gilman SE, Striegel-Moore RH. Validity and reliability of a fujian translation and adaptation of the eating disorder examination questionnaire. *The International Journal of Eating Disorders*. 2010; 43(2):171–178. [PubMed: 19308995]
- Berg, KC.; Peterson, CB.; Frazier, P.; Crow, SC. Psychometric evaluation of the Eating Disorder Examination and Eating Disorder Examination-Questionnaire: A systematic review of the literature. 2010. Manuscript accepted for publication
- *. Binford RB, Le Grange D, Jellar CC. Eating disorder examination versus eating disorder examination-questionnaire in adolescents with full and partial syndrome bulimia nervosa and anorexia nervosa. *International Journal of Eating Disorders*. 2005; 37:44–49. [PubMed: 15690465]
- *. Black CMD, Wilson GT. Assessment of eating disorders: Interview versus questionnaire. *The International Journal of Eating Disorders*. 1996; 20(1):43–50. [PubMed: 8807351]
- Bryant-Waugh RJ, Cooper PJ, Taylor CL, Lask BD. The use of the eating disorder examination with children: A pilot study. *The International Journal of Eating Disorders*. 1996; 19(4):391–397. [PubMed: 8859397]
- Butcher, JN.; Dahlstrom, WG.; Graham, JR.; Tellegen, A.; Kraemmer, B. MMPI-2: Manual for administration and scoring. University of Minnesota Press; Minneapolis, MN: 1989.
- Byrne SM, Allen KL, Lampard AM, Dove ER, Fursland A. The factor structure of the eating disorder examination in clinical and community samples. *The International Journal of Eating Disorders*. 2010; 43(3):260–265. [PubMed: 19350647]
- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*. 1959; 56:81–105. [PubMed: 13634291]
- Campbell R, Quilty B, Dieppe P. Discrepancies between patient's assessments of outcome: Qualitative study nested within a randomised controlled trial. *British Medical Journal*. 2003; 326:252–253. [PubMed: 12560274]
- *. Carter JC, Aime AA, Mills JS. Assessment of bulimia nervosa: A comparison of interview and self-report questionnaire methods. *International Journal of Eating Disorders*. 2001; 30:187–192. [PubMed: 11449452]
- Celio AA, Wilfley DE, Crow SJ, Mitchell J, Walsh BT. A comparison of the binge eating scale, questionnaire for eating and weight patterns-revised, and eating disorder examination questionnaire with instructions with the eating disorder examination in the assessment of binge eating disorder and its symptoms. *The International Journal of Eating Disorders*. 2004; 36(4):434–444. [PubMed: 15558644]
- Cohen, J. Statistical power analysis for the behavioral sciences. 2nd ed.. Erlbaum; Hillsdale, NJ: 1988.
- Cooper Z, Cooper PJ, Fairburn CG. The validity of the eating disorder examination and its subscales. *The British Journal of Psychiatry*. 1989; 154(6):807–812. [PubMed: 2597887]
- *. de Zwaan M, Mitchell JE, Swan-Kremeier L, McGregor T, Howell ML, Roerig JL, Crosby RD. A comparison of different methods of assessing the features of eating disorders in post-gastric bypass patients: A pilot study. *European Eating Disorders Review*. 2004; 12:380–386.
- Fairburn, CG.; Beglin, S. Eating disorder examination questionnaire.. In: Fairburn, CG., editor. *Cognitive behavior therapy and eating disorders* (). Guilford Press; New York: 2008.

- *. Fairburn CG, Beglin SJ. Assessment of eating disorders: Interview or self-report questionnaire? *The International Journal of Eating Disorders*. 1994; 16(4):363–370. [PubMed: 7866415]
- Fairburn, CG.; Cooper, Z. The eating disorder examination (12th edition).. In: Fairburn, CG.; Wilson, GT., editors. *Binge eating: Nature, assessment, and treatment*. Guilford Press; New York: 1993. p. 317-360.
- Fairburn CG, Cooper Z, Doll HA, Norman P, O'Connor M. The natural course of bulimia nervosa and binge eating disorder in young women. *Archives of General Psychiatry*. 2000; 57:659–665. [PubMed: 10891036]
- Fairburn, CG.; Cooper, Z.; O'Connor, M. Eating disorder examination, edition 16.0D.. In: Fairburn, CG., editor. *Cognitive behavior therapy and eating disorders ()*. Guilford Press; New York: 2008.
- *. Goldfein JA, Devlin MJ, Kamenetz C. Eating disorder examination-questionnaire with and without instruction to assess binge eating in patients with binge eating disorder. *The International Journal of Eating Disorders*. 2005; 37(2):107–111. [PubMed: 15732081]
- Grilo, CM. Structured Instruments.. In: Mitchell, JE.; Peterson, CB., editors. *Assessment of Eating Disorders*. The Guilford Press; New York, NY: 2005.
- Grilo CM, Crosby RD, Peterson CB, Masheb RM, White MA, Crow SJ, Wonderlich SA, Mitchell JE. Factor structure of the eating disorder examination interview in patients with binge-eating disorder. *Obesity*. 2009; 18:977–981. [PubMed: 19798064]
- Grilo CM, Lozano C, Elder KA. Inter-rater and test-retest reliability of the spanish language version of the eating disorder examination interview: Clinical and research implications. *Journal of Psychiatric Practice*. 2005; 11(4):231–240. [PubMed: 16041233]
- Grilo CM, Masheb RM, Lozano-Blanco C, Barry DT. Reliability of the eating disorder examination in patients with binge eating disorder. *The International Journal of Eating Disorders*. 2004; 35(1):80–85. [PubMed: 14705160]
- *. Grilo CM, Masheb RM, Wilson GT. A comparison of different methods for assessing the features of eating disorders in patients with binge eating disorder. *Journal of Consulting and Clinical Psychology*. 2001a; 69(2):317–322. [PubMed: 11393608]
- *. Grilo CM, Masheb RM, Wilson GT. Different methods for assessing the features of eating disorders in patients with binge eating disorder: A replication. *Obesity Research*. 2001b; 9:418–422. [PubMed: 11445665]
- Higgins J, Thompson S. Quantifying heterogeneity in meta-analysis. *Statistics in Medicine*. 2002; 21:1539–1558. [PubMed: 12111919]
- Hoek HW. Incidence, prevalence, and mortality of anorexia and other eating disorders. *Current Opinion in Psychiatry*. 2006; 19:389–394. [PubMed: 16721169]
- *. Kalarchian MA, Wilson GT, Brolin RE, Bradley L. Assessment of eating disorders in bariatric surgery candidates: Self-report questionnaire versus interview. *The International Journal of Eating Disorders*. 2000; 28(4):465–469. [PubMed: 11054796]
- Keel PK, Crow SJ, Davis TL, Mitchell JE. Assessment of eating disorders: Comparison of interview and questionnaire data from a long-term follow-up study of bulimia nervosa. *Journal of Psychosomatic Research*. 2002; 53:1043–1047. [PubMed: 12445594]
- Lavender JM, Anderson DA. Effect of perceived anonymity in assessments of eating disordered behaviors and attitudes. *International Journal of Eating Disorders*. 2009; 42:546–551. [PubMed: 19172594]
- Le Grange D, Crosby RD, Rathouz PJ, Leventhal BL. A randomized controlled comparison of family-based treatment and supportive psychotherapy for adolescent bulimia nervosa. *Archives of General Psychiatry*. 2007; 64:1049–1056. [PubMed: 17768270]
- Lipsey, MW.; Wilson, DB. *Practical meta-analysis*. Sage Publications, Inc.; Thousand Oaks, California: 2001.
- Loeb KL, Pike KM, Walsh BT, Wilson GT. Assessment of diagnostic features of bulimia nervosa: Interview versus self-report format. *The International Journal of Eating Disorders*. 1994; 16(1):75–81. [PubMed: 7920584]
- Mond JM, Hay PJ, Rodger B, Owen C. Self-report versus interview assessment of purging in a community sample of women. *European Eating Disorders Review*. 2007; 15(6):403–409. [PubMed: 17960776]

- *. 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. *Behaviour Research and Therapy*. 2004; 42(5):551–567. [PubMed: 15033501]
- *. Passi VA, Bryson SW, Lock J. Assessment of eating disorders in adolescents with anorexia nervosa: Self-report questionnaire versus interview. *The International Journal of Eating Disorders*. 2003; 33(1):45–54. [PubMed: 12474198]
- Peterson CB, Crosby RD, Wonderlich SA, Joiner T, Crow SJ, Mitchell JE, Bardone-Cone AM, Klein M, Le Grange D. Psychometric properties of the eating disorder examination-questionnaire: Factor structure and internal consistency. *The International Journal of Eating Disorders*. 2007; 40(4):386–389. [PubMed: 17304585]
- Peterson CB, Mitchell JE, Crow SJ, Crosby RD, Wonderlich SA. The efficacy of self-help group treatment and therapist-led group treatment for binge eating disorder. *American Journal of Psychiatry*. 2009; 166:1347–1354. [PubMed: 19884223]
- Pretorius N, Waller G, Gowers S, Schmidt U. Validity of the eating disorders examination-questionnaire when used with adolescents with bulimia nervosa and atypical bulimia nervosa. *Eating and Weight Disorders*. 2009; 14:e243–e248. [PubMed: 20179414]
- Reas DL, Grilo CM, Masheb RM. Reliability of the eating disorder examination-questionnaire in patients with binge eating disorder. *Behaviour Research and Therapy*. 2006; 44(1):43–51. [PubMed: 16301013]
- Reeves JC, Large RG, Honeyman M. Parasuicide and depression: A comparison of clinical and questionnaire diagnoses. *Australian and New Zealand Journal of Psychiatry*. 1985; 19:30–33. [PubMed: 3859282]
- Sellbom M, Ben-Porath YS, McNulty JL, Arbisi PA, Graham JT. Elevation differences between MMPI-2 clinical and restructured clinical (RC) scales: Frequency, origins, and interpretative implications. *Assessment*. 2006; 13:430–441. [PubMed: 17050913]
- *. Sysko R, Walsh BT, Fairburn CG. Eating disorder examination-questionnaire as a measure of change in patients with bulimia nervosa. *The International Journal of Eating Disorders*. 2005; 37(2):100–106. [PubMed: 15732070]
- *. Sysko R, Walsh BT, Schebendach J, Wilson GT. Eating behavior among women with anorexia nervosa. *The American Journal of Clinical Nutrition*. 2005; 82:296–301. [PubMed: 16087971]
- *. Wilfley DE, Schwartz MB, Spurrell EB, Fairburn CG. Assessing the specific psychopathology of binge eating disorder patients: Interview or self-report? *Behaviour Research and Therapy*. 1997; 35(12):1151–1159. [PubMed: 9465449]
- Wilson, GT. Assessment of binge eating.. In: Fairburn, CG.; Wilson, GT., editors. *Binge eating: Nature, assessment, and treatment*. Guilford Press; New York: 1993.
- *. Wolk SL, Loeb KL, Walsh BT. Assessment of patients with anorexia nervosa: Interview or self-report. *International Journal of Eating Disorders*. 2005; 37:92–99. [PubMed: 15732073]
- Woodruff-Borden J, Jeffery SE, Bourland SL, Brothers AJ, Albano AM. Patient self-report in the assessment of panic disorders: Comparison with interview-derived clinician ratings. *Journal of Nervous and Mental Disease*. 2000; 188:308–310. [PubMed: 10830570]
- Zimmerman M, Coryell WH. Diagnosing personality disorders in the community: A comparison of self-report and interview measures. *Archives of General Psychiatry*. 1990; 47:527–531. [PubMed: 2350205]

Table 1

Sample Characteristics of the 16 Studies Used in the Meta-analysis

Studies	Diagnostic Group	N	Mean Age	Gender	Context
Barnes et al. (in press)	BED	66	44.1	76% Female	Treatment Study
Binford et al. (2005)	AN	21	15.8	96% Female	Treatment Study
	BN	25			
	EDNOS	24			
Black & Wilson (1996)	SA	48	32.6	100% Female	Clinic Treatment
Carter et al. (2001)	BN	60	26.0	100% Female	Clinic Treatment
de Zwaan et al. (2004)	BS	45	42.0	87% Female	Clinic Treatment
Fairburn & Beglin (1994)	CS	285	26.6	100% Female	Epidemiological Study
	AN, BN	13, 23	15.8, 24.3	100% Female	Clinic Treatment
Goldfein et al. (2005)	BED	37	>18.0	Male & Female	Treatment Study
Grilo et al. (2001) ^a	BED	82	42.1	82% Female	Treatment Study
Grilo et al. (2001) ^b	BED	47	46.2	77% Female	Treatment Study
Kalarchian et al. (2000)	BS	98	38.4	78% Female	Clinic Treatment
Mond et al. (2004)	CS	208	35.4	100% Female	Epidemiological Study
Passi et al. (2002)	AN	28	15.1	100% Female	Treatment Study
Sysko, Walsh, Schebendach et al. (2005)	AN	12	21.3	100% Female	Clinic Treatment
Sysko, Walsh, & Fairburn (2005)	BN	50	31.7	100% Female	Treatment study
Wilfley et al. (1997)	BED	52	45.4	89% Female	Treatment Study
Wolk et al. (2005)	AN	60	24.7	100% Female	Clinic Treatment

Note: BED = Binge Eating Disorder; AN = Anorexia Nervosa; BN = Bulimia Nervosa; EDNOS = Eating Disorder Not Otherwise Specified; SA = Substance Abuse Sample; BS = Bariatric Surgery Patients; CS = Community Sample

Table 2
Meta-Analytic Findings of the Convergence of Subscale Scores Using Correlation Coefficients

Studies	Restrained ($Q(14) = 12.58$) ($I^2 = 0.00\%$)			Eating Concern ($Q(11) = 16.40$) ($I^2 = 32.93\%$)			Shape Concern ($Q(15) = 16.59$) ($I^2 = 29.57\%$)			Weight Concern ($Q(15) = 18.59$) ($I^2 = 19.33\%$)		
	ES	LL	UL	ES	LL	UL	ES	LL	UL	ES	LL	UL
Anorexia Nervosa												
Binford et al. (2005)	.71			.90			.89			.83		
Passi et al. (2003)	.71			.67			.91			.82		
Wolk et al. (2005)	.49			.51			.83			.61		
<i>Group Effect Size Estimates</i>	.62	.43	.76	.72	.49	.86	.88	.76	.94	.75	.60	.85
Bulimia Nervosa												
Binford et al. (2005)	.79			.75			.85			.87		
Carter et al. (2001)	-			-			.43			.54		
<i>Group Effect Size Estimates</i>	.79	.47	.93	.75	.26	.93	.66	.33	.84	.70	.48	.83
Combined AN & BN Sample												
Fairburn & Beglin (1994)	.78			-			.83			.85		
<i>Group Effect Size Estimates</i>	.78	.53	.91	-	-	-	.83	.51	.95	.85	.66	.94
EDNOS												
Binford et al. (2005)	.85			.94			.82			.88		
<i>Group Effect Size Estimates</i>	.85	.63	.94	.94	.79	.98	.82	.45	.95	.88	.69	.96
Binge Eating Disorder												
Barnes et al. (2010)	.57			.53			.54			.57		
Grilo et al. (2001)a	.69			.33			.56			.66		
Grilo et al. (2001)b	.59			.55			.42			.63		
Wilfley et al. (1997)	.66			.59			.69			.63		
<i>Group Effect Size Estimates</i>	.63	.49	.74	.50	.25	.69	.56	.33	.73	.62	.48	.73
Community Sample												
Fairburn & Beglin (1994)	.81			-			.80			.79		
Mond et al. (2004)b	.71			.68			.78			.77		
<i>Group Effect Size Estimates</i>	.77	.65	.85	.68	.28	.88	.79	.60	.90	.78	.66	.86
Bariatric Surgery Patients												

Studies	Restraint ($Q(14) = 12.58$) ($I^2 = 0.00\%$)			Eating Concern ($Q(11) = 16.40$) ($I^2 = 32.93\%$)			Shape Concern ($Q(15) = 16.59$) ($I^2 = 29.57\%$)			Weight Concern ($Q(15) = 18.59$) ($I^2 = 19.33\%$)		
	95% CI			95% CI			95% CI			95% CI		
	ES	LL	UL	ES	LL	UL	ES	LL	UL	ES	LL	UL
de Zwaan et al. (2004)	.54			.80			.80			.79		
Kalarchian et al. (2000)	.60			.62			.77			.71		
<i>Group Effect Size Estimates</i>	<i>.58</i>	<i>.36</i>	<i>.73</i>	<i>.71</i>	<i>.45</i>	<i>.86</i>	<i>.78</i>	<i>.57</i>	<i>.90</i>	<i>.75</i>	<i>.59</i>	<i>.85</i>
Substance Abuse Patients												
Black & Wilson (1996)	.75			-			.84			.85		
<i>Group Effect Size Estimates</i>	<i>.75</i>	<i>.50</i>	<i>.89</i>	-	-	-	<i>.84</i>	<i>.55</i>	<i>.95</i>	<i>.85</i>	<i>.68</i>	<i>.93</i>
Overall Effect Size Estimates	.69	.63	.74	.68	.58	.76	.76	.69	.81	.74	.69	.79

Note. Effect size estimates for specific diagnostic groups are in italics. Overall effect size estimates are in boldface. CI = confidence interval; ES = effect size; LL = Lower limit; UL = upper limit; AN = Anorexia Nervosa; BN = Bulimia Nervosa; EDNOS = Eating Disorder Not Otherwise Specified.

Table 3

Meta-Analytic Findings of the Convergence of Subscale Scores Using Cohen's d

Studies	Restraint ($Q(14) = 15.96$) ($I^2 = 12.26\%$)			Eating Concern ($Q(14) = 19.93$) ($I^2 = 29.75\%$)			Shape Concern ($Q(15) = 19.35$) ($I^2 = 22.47\%$)			Weight Concern ($Q(15) = 16.84$) ($I^2 = 10.91\%$)		
	ES	95% CI		ES	95% CI		ES	95% CI		ES	95% CI	
		LL	UL		LL	UL		LL	UL		LL	UL
Anorexia Nervosa												
Binford et al. (2005)	-0.09			0.18			0.16			0.14		
Passi et al. (2003)	0.25			0.68			0.36			0.25		
Sysko, Walsh, Schebendach, et al. (2005)	0.60			0.29			0.61			0.62		
Wolk et al. (2005)	0.27			0.50			0.32			0.45		
Group Effect Size Estimates	0.22	-0.03	0.47	0.43	0.05	0.82	0.34	-0.11	0.80	0.35	0.04	0.66
Bulimia Nervosa												
Binford et al. (2005)	0.35			0.87			0.44			0.26		
Carter et al. (2001)	-			-			-0.21			-0.25		
Group Effect Size Estimates	0.35	-0.26	0.96	0.87	0.04	1.70	0.07	-0.55	0.68	-0.08	-0.48	0.80
EDNOS												
Binford et al. (2005)	0.05			0.11			0.23			0.23		
Group Effect Size Estimates	0.05	-0.51	0.60	0.11	-0.66	0.88	0.23	-0.68	1.15	0.23	-0.41	0.87
Binge Eating Disorder												
Barnes et al. (2010)	0.30			0.71			0.88			0.52		
Grilo et al. (2001)a	0.66			0.46			1.06			0.73		
Grilo et al. (2001)b	0.46			1.76			1.72			0.65		
Wilfley et al. (1997)	0.37			1.35			1.00			0.67		
Group Effect Size Estimates	0.46	0.28	0.64	1.01	0.68	1.34	1.10	0.73	1.56	0.64	0.40	0.89
Community Sample												
Fairburn & Beglin (1994)	0.26			0.47			0.59			0.35		
Mond et al. (2004)b	0.19			0.53			0.65			0.44		
Group Effect Size Estimates	0.23	0.09	0.36	0.50	0.10	0.91	0.62	0.09	1.15	0.39	0.13	0.66
Bariatric Surgery Patients												
de Zwaan et al. (2004)	0.82			0.33			0.00			0.00		
Kalarchian et al. (2000)	0.33			0.75			0.74			0.65		

Studies	Restraint ($Q(14) = 15.96$) ($I^2 = 12.26\%$)			Eating Concern ($Q(14) = 19.93$) ($I^2 = 29.75\%$)			Shape Concern ($Q(15) = 19.35$) ($I^2 = 22.47\%$)			Weight Concern ($Q(15) = 16.84$) ($I^2 = 10.91\%$)		
	95% CI			95% CI			95% CI			95% CI		
	ES	LL	UL	ES	LL	UL	ES	LL	UL	ES	LL	UL
<i>Group Effect Size Estimates</i>												
Substance Abuse Patients	0.48	0.24	0.71	0.55	0.10	1.00	0.50	-0.17	0.97	0.39	0.06	0.72
Black & Wilson (1996)	0.20			0.36			-0.06			0.21		
<i>Group Effect Size Estimates</i>	0.20	-0.21	0.60	0.36	-0.05	0.76	-0.06	-0.88	0.77	0.21	-0.30	0.72
Overall Effect Size Estimates	0.31	0.22	0.40	0.62	0.45	0.80	0.54	0.33	0.75	0.39	0.26	0.51

Note. Effect size estimates for specific diagnostic groups are in italics. Overall effect size estimates are in boldface. CI = confidence interval; ES = effect size; LL = Lower limit; UL = upper limit; EDNOS = Eating Disorder Not Otherwise Specified.

Table 4
Meta-Analytic Findings of the Convergence of Behavior Frequency Scores Using Correlation Coefficients

Studies	Objective Bulimic Episodes (Q (14) = 9.22) (I^2 = 0.00%)			Subjective Bulimic Episodes (Q (8) = 4.74) (I^2 = 0.00%)			Self-Induced Vomiting (Q (7) = 8.78) (I^2 = 20.31%)			Laxative Misuse (Q (5) = 4.58) (I^2 = 0.00%)		
	ES	LL	UL	ES	LL	UL	ES	LL	UL	ES	LL	UL
Anorexia Nervosa												
Wolk et al. (2005)	.92			-			.88			.70		
<i>Group Effect Size Estimates</i>	.92	.47	.99	-	-	-	.88	.42	.98	.70	-.71	.99
Bulimia Nervosa												
Binford et al. (2005)	.48			.21			.73			-		
Carter et al. (2001)	.56			.46			.72			.88		
Sysko, Walsh, & Fairburn (2005)	.63			.60			.88			.99		
<i>Group Effect Size Estimates</i>	.56	.00	.85	.44	-.08	.77	.79	.48	.93	.96	.65	1.00
Combined AN & BN Sample												
Fairburn & Beglin (1994)	.60			-			.91			.89		
<i>Group Effect Size Estimates</i>	.60	-.39	.95	-	-	-	.91	.52	.99	.89	-.33	1.00
EDNOS												
Binford et al. (2005)	.40			.50			.93			-		
<i>Group Effect Size Estimates</i>	.40	-.61	.91	.50	-.41	.91	.93	.59	.99	-	-	-
Binge Eating Disorder												
Barnes et al. (2010)	.34			.21			-			-		
Goldfein et al. (2005)	.20			-			-			-		
Grilo et al. (2001)a	.29			-.06			-			-		
Grilo et al. (2001)b	.28			-.09			-			-		
Wilfley et al. (1997)	.20			-			-			-		
<i>Group Effect Size Estimates</i>	.26	-.21	.64	.02	-.47	.51	-	-	-	-	-	-
Community Sample												
Fairburn & Beglin (1994)	.45			-			.88			.60		
Mond et al. (2004)b	.93			.78			-			-		
<i>Group Effect Size Estimates</i>	.79	.31	.95	.78	.14	.96	.88	.44	.98	.60	-.78	.98
Bariatric Surgery Patients												

Studies	Objective Bulimic Episodes (Q (14) = 9.22) (I^2 = 0.00%)			Subjective Bulimic Episodes (Q (8) = 4.74) (I^2 = 0.00%)			Self-Induced Vomiting (Q (7) = 8.78) (I^2 = 20.31%)			Laxative Misuse (Q (5) = 4.58) (I^2 = 0.00%)		
	ES	LL	UL	ES	LL	UL	ES	LL	UL	ES	LL	UL
de Zwaan et al. (2004)	-			.41			-			-		
Kalarchian et al. (2000)	.46			-			-			-		
<i>Group Effect Size Estimates</i>	<i>.46</i>	<i>.29</i>	<i>.60</i>	<i>.41</i>	<i>-.47</i>	<i>.88</i>	-	-	-	-	-	-
Substance Abuse Patients												
Black & Wilson (1996)	.53			-			.99			.99		
<i>Group Effect Size Estimates</i>	<i>.53</i>	<i>-.46</i>	<i>.93</i>	-	-	-	<i>.99</i>	<i>.94</i>	<i>1.00</i>	<i>.99</i>	<i>.71</i>	<i>1.00</i>
Overall Effect Size Estimates	.55	.32	.71	.37	.07	.61	.90	.82	.95	.92	.71	.98

Note. Effect size estimates for specific diagnostic groups are in italics. Overall effect size estimates are in boldface. CI = confidence interval; ES = effect size; LL = Lower limit; UL = upper limit; AN = Anorexia Nervosa; BN = Bulimia Nervosa; EDNOS = Eating Disorder Not Otherwise Specified.

Table 5
Meta-Analytic Findings of the Convergence of Behavior Frequency Scores Using Cohen's d

Studies	Objective Bulimic Episodes ($Q(13) = 10.25$) ($I^2 = 0.00\%$)			Subjective Bulimic Episodes ($Q(8) = 9.87$) ($I^2 = 18.96\%$)		
	ES	LL	UL	ES	LL	UL
Anorexia Nervosa						
Wolk et al. (2005)	0.16			-		
<i>Group Effect Size Estimates</i>	0.16	-0.49	0.78	-	-	-
Bulimia Nervosa						
Binford et al. (2005)	-0.58			-0.25		
Carter et al. (2001)	-0.16			-0.26		
Sysko, Walsh, & Fairburn (2005)	-0.39			-0.53		
<i>Group Effect Size Estimates</i>	-0.35	-0.75	0.06	-0.36	-0.63	-0.09
EDNOS						
Binford et al. (2005)	0.07			-0.57		
<i>Group Effect Size Estimates</i>	0.07	-0.69	0.84	-0.57	-1.16	0.02
Binge Eating Disorder						
Barnes et al. (2010)	-0.36			-0.33		
Goldfein et al. (2005)	0.24			-		
Grilo et al. (2001)a	-0.22			-0.06		
Grilo et al. (2001)b	-0.31			0.12		
Wilfley et al. (1997)	-0.51			-		
<i>Group Effect Size Estimates</i>	-0.24	-0.54	0.05	-0.11	-0.33	0.12
Community Sample						
Fairburn & Beglin (1994)	0.26			-		
Mond et al. (2004)b	-0.50			-0.30		
<i>Group Effect Size Estimates</i>	-0.11	-0.51	0.29	-0.30	-0.57	-0.03
Bariatric Surgery Patients						
Kalarchian et al. (2000)	-0.22			0.17		
<i>Group Effect Size Estimates</i>	-0.22	-0.82	0.39	0.17	-0.28	0.62
Substance Abuse Patients						

Studies	Objective Bulimic Episodes (<i>Q</i> (13) = 10.25) (<i>I</i> ² = 0.00%)			Subjective Bulimic Episodes (<i>Q</i> (8) = 9.87) (<i>I</i> ² = 18.96%)		
	95% CI			95% CI		
	<i>ES</i>	<i>LL</i>	<i>UL</i>	<i>ES</i>	<i>LL</i>	<i>UL</i>
Black & Wilson (1996)	0.16			-		
Group Effect Size Estimates	0.16	-0.50	0.83	-	-	-
Overall Effect Size Estimates	-0.16	-0.34	0.01	-0.22	-0.35	-0.08

Note. CI = confidence interval; *ES* = effect size; *LL* = Lower limit; *UL* = upper limit; EDNOS: Eating Disorder Not Otherwise Specified.