

Development and Validation of the Eating Disorder Diagnostic Scale: A Brief Self-Report Measure of Anorexia, Bulimia, and Binge-Eating Disorder

Eric Stice
University of Texas at Austin

Christy F. Telch
Stanford University

Shireen L. Rizvi
University of Washington

This article describes the development and validation of a brief self-report scale for diagnosing anorexia nervosa, bulimia nervosa, and binge-eating disorder. Study 1 used a panel of eating-disorder experts and provided evidence for the content validity of this scale. Study 2 used data from female participants with and without eating disorders ($N = 367$) and suggested that the diagnoses from this scale possessed temporal reliability (mean $\kappa = .80$) and criterion validity (with interview diagnoses; mean $\kappa = .83$). In support of convergent validity, individuals with eating disorders identified by this scale showed elevations on validated measures of eating disturbances. The overall symptom composite also showed test-retest reliability ($r = .87$), internal consistency (mean $\alpha = .89$), and convergent validity with extant eating-pathology scales. Results implied that this scale was reliable and valid in this investigation and that it may be useful for clinical and research applications.

It has been estimated that 10% of female individuals in western countries will suffer from a diagnosable eating disorder (American Psychiatric Association [APA], 1994), making it one of the more prevalent psychiatric problems faced by women. Anorexia nervosa is characterized by (a) extreme emaciation; (b) intense fear of gaining weight or becoming fat despite a low body weight; (c) disturbed perception of weight and shape, an undue influence of weight or shape on self-evaluation, or a denial of the seriousness of the low body weight; and (d) amenorrhea (APA, 1994). This disorder has a lifetime prevalence of almost 1% among females, is refractory to treatment, shows a chronic course, results in serious medical complications, and is associated with psychiatric comorbidity such as mood, anxiety, and personality disorders (Wilson, Heffernan, & Black, 1996).

Bulimia nervosa involves (a) recurrent episodes of uncontrollable consumption of large amounts of food, (b) compensatory

behavior to prevent consequent weight gain (e.g., vomiting, laxative abuse, diuretic abuse, or excessive exercise), and (c) undue influence of weight and shape on self-evaluation (APA, 1994). The lifetime prevalence for bulimia nervosa is approximately 2% for females (Newman et al., 1996; Whitaker et al., 1990). This disorder is marked by a persistent course and is associated with high levels of comorbid psychopathology, including affective disorders, anxiety disorders, and substance abuse (Garfinkel et al., 1995; Keller, Herzog, Lavori, Bradburn, & Mahoney, 1992).

Binge-eating disorder involves (a) repeated episodes of uncontrollable binge eating characterized by certain features (e.g., rapid eating or eating alone because of embarrassment), (b) marked distress regarding binge eating, and (c) the absence of compensatory behaviors (APA, 1994). The lifetime prevalence for binge-eating disorder is approximately 4% in the community (Spitzer et al., 1993), but it has been estimated that about 30% of individuals presenting for weight-control treatment meet criteria for this disorder (Brody, Walsh, & Devlin, 1994). This disorder appears to have a persistent course and is associated with obesity, weight cycling, health complications, and psychiatric comorbidity (Wilson et al., 1996; Telch & Stice, 1998).

Research on the etiology, prevention, and treatment of eating disorders has increased dramatically over the past 2 decades (Smolak, Levine, & Striegel-Moore, 1996), but progress has been limited in part by the scarcity of validated measures of eating disorders. Although there are structured psychiatric interviews for arriving at *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV*; APA, 1994) diagnoses of anorexia nervosa, bulimia nervosa, and binge-eating disorder (e.g., the Eating Disorder Examination [EDE; Fairburn & Cooper, 1993] and the Structured Clinical Interview for *DSM* [SCID; Spitzer, Williams,

Eric Stice, Department of Psychology, University of Texas at Austin; Christy F. Telch, Department of Psychiatry and Behavioral Sciences, Stanford University; Shireen L. Rizvi, Department of Psychology, University of Washington.

This study was supported by Postdoctoral Fellowship MH19908, Career Award MH01708, Exploratory Development Grant MH54641, and Research Grant MH50271 from the National Institute of Mental Health, as well as by a grant from the McKnight Foundation. We thank Kim Fromme, Helen C. Kraemer, Heather Shaw, and Michael J. Telch for their thoughtful input regarding this article.

Correspondence concerning this article should be addressed to Eric Stice, Department of Psychology, 330 Mezes Hall, University of Texas at Austin, Austin, Texas 78712. Electronic mail may be sent to stice@psy.utexas.edu.

Gibbon, & First, 1990)), no parallel self-report questionnaire measure exists. Available self-report questionnaires, such as the Bulimia Test—Revised (Thelen, Farmer, Wonderlich, & Smith, 1991), Questionnaire on Eating and Weight Patterns (Spitzer et al., 1992), Eating Disorder Examination—Questionnaire (Fairburn & Beglin, 1994), and Questionnaire for Eating Disorder Diagnoses (Mintz, O'Halloran, Mulholland, & Schneider, 1997), do not yield *DSM-IV* diagnoses for all three eating disorders. Moreover, the reliability and validity of measures from extant self-report questionnaires have not been well established. For example, we were unable to locate any published studies that report the test-retest reliability for *DSM-IV* eating-disorder diagnoses generated by these self-report questionnaires. Similarly, there appear to be no published studies that report the agreement between *DSM-IV* eating-disorder diagnoses ascertained through validated structured interviews and those generated by these self-report questionnaires.

A brief self-report questionnaire would be very useful for etiologic research, in that it is time consuming and expensive to conduct lengthy structured psychiatric interviews with large numbers of participants. Such a questionnaire might also be useful when researchers need relatively frequent measurement of eating pathology (e.g., prevention programs or treatment studies), as it would minimize the response burden for participants. Finally, such a measure might prove useful in clinical settings (e.g., primary care offices) where a brief questionnaire is desired to identify individuals with eating pathology.

Accordingly, we sought to develop a brief self-report scale that would provide diagnoses of anorexia nervosa, bulimia nervosa, and binge-eating disorder, and to generate evidence of the reliability and validity of this new diagnostic instrument. Study 1 describes the processes by which we generated items and examined the content validity of this new scale. Study 2 assesses the test-retest reliability, criterion validity, and convergent validity of this diagnostic scale. As some researchers may be interested in using an overall eating-disorder symptom composite for etiologic, prevention, or treatment studies (because of the low base rate of these disorders), a tertiary aim was to explore the psychometric properties of such a symptom composite derived from this diagnostic scale. In particular, the test-retest reliability, internal consistency, and convergent validity of the overall symptom composite were examined.

Study 1

Overview

The purpose of Study 1 was to (a) generate a set of items to assess the *DSM-IV* diagnostic criteria for anorexia nervosa, bulimia nervosa, and binge-eating disorder for an Eating Disorder Diagnostic Scale (EDDS), (b) document that all relevant symptoms were included and that no irrelevant ones were included, and (c) refine item content and instructions. Following recommendations regarding content validation (Haynes, Richard, & Kubany, 1995), we used a multistep process to accomplish these aims.

Method

Step 1

Items were derived from multiple sources to enhance the content validity of the scale. Items assessing *DSM-IV* diagnostic criteria for anorexia

nervosa, bulimia nervosa, and binge-eating disorder were adapted from validated structured psychiatric interviews, which assessed these disorders: the EDE (Fairburn & Cooper, 1993) and the eating-disorder module of the SCID (Spitzer et al., 1990). We used the version of the EDE that assesses diagnostic criteria for anorexia nervosa, bulimia nervosa, and binge-eating disorder. We also consulted the *DSM-IV* (APA, 1994) regarding diagnostic criteria for these three eating disorders in an effort to ensure that all relevant symptoms were represented on the scale. Items were worded to capture diagnostic criteria for these disorders, including the necessary time windows for various behaviors.

Step 2

To document that all of the diagnostic criteria for the three eating disorders were included and that no irrelevant information was included in the EDDS, a panel of eating-disorders researchers from around the country evaluated a preliminary version of this questionnaire. Twenty-six eating-disorder experts, identified through a literature search, were mailed the preliminary draft of the EDDS, a list of the diagnostic criteria taken directly from the *DSM-IV*, and a cover letter with instructions. Experts were asked to (a) check that all of the *DSM-IV* diagnostic criteria were assessed on EDDS, (b) cross out any EDDS items that did not reflect a *DSM-IV* symptom, (c) suggest refinements to the wording of items or instructions, and (d) write down any additional items that should be included. Of the 26 participants who were mailed this information, 14 (54%) returned completed packets. Four of the 14 experts noted that one diagnostic criteria was not reflected on the EDDS (fasting) and made suggestions regarding wording of this item. Experts agreed that all remaining diagnostic criteria were included and that no irrelevant items were included. Experts also provided several useful suggestions that clarified the wording of items and instructions. These responses were used to guide a revision of the preliminary EDDS.

Step 3

The revised draft of EDDS was pilot tested with patients from an eating-disorders clinic, undergraduate students, and high school students ($N = 15$). These participants were asked to comment on the clarity of the questions and instructions. This feedback was used to guide a final revision of the preliminary EDDS, which primarily centered on clarifying the instructions.

Results

The above item-generation and refinement procedures resulted in a 22-item version of the EDDS. Items used a combination of Likert, yes-no, frequency, and write-in response formats to assess all of the *DSM-IV* diagnostic symptoms for anorexia nervosa, bulimia nervosa, and binge-eating disorder. In an effort to minimize the effects of idiosyncratic conceptions of what constitutes a "binge," we did not use this term in the scale. The EDDS is included in the Appendixes, along with the scoring algorithm:

Study 2

Overview

The purpose of Study 2 was to examine the test-retest reliability, criterion validity, and convergent validity of the EDDS eating-disorder diagnoses, as well as the test-retest reliability, internal consistency, and convergent validity of the EDDS symptom composite.

Method

Participants

An effort was made to recruit a heterogeneous sample that was diverse in age, socioeconomic status, and geographic location to maximize the generalizability of the findings. Participants were 367 females between the ages of 13 and 65 recruited from the metropolitan areas of San Francisco, New York, Minneapolis-St. Paul, and Austin. Participants were recruited from several ongoing projects, including (a) a randomized clinical trial for the treatment of anorexia nervosa ($n = 12$), (b) a randomized clinical trial for the treatment of bulimia nervosa ($n = 3$), (c) a randomized clinical trial for the treatment of binge-eating disorder ($n = 17$), (d) a multisite study on the longitudinal course of threshold and subthreshold eating disorders in a non-treatment seeking sample ($n = 185$), (e) a multisite study on affect ($n = 38$), (f) a longitudinal study of the risk factors for eating disorders ($n = 109$), and (g) an inpatient psychiatric treatment unit ($n = 3$).

Participants from the first six sources were recruited directly from the community through advertisements placed in local media, fliers distributed at universities and medical clinics, and direct mailings to eligible females between the ages of 13 and 65 to participate in these research studies. The 3 patients recruited from the inpatient treatment unit were approached directly because a chart review suggested that they had a diagnosis of anorexia nervosa. Data for this study were collected at: (a) various assessment points in the three randomized clinical trials and the longitudinal course study, (b) at the baseline assessment of the affect study and the longitudinal risk factor study, and (c) within the first month of treatment for the 3 anorexic individuals from the inpatient treatment unit.

Participants ranged in age from 13 to 61 ($M = 29.7$, $SD = 13.2$). The sample was composed of participants who were 2% Asian or Pacific Islander, 2% Black, 6% Hispanic, 1% Native American, 80% Caucasian, and 9% who specified "other" or mixed-racial heritage. Educational attainment ranged from some high school (22%) to graduate or professional degree (17%), with a mode of some college education (29%).

Measures

Structured Psychiatric Interview. The EDE (Fairburn & Cooper, 1993) was the primary "gold standard" against which our new self-report diagnostic scale was compared for validation purposes. The EDE is a structured psychiatric interview that assesses diagnostic criteria for *DSM-IV* eating disorders. We used the version that assesses diagnostic criteria for anorexia nervosa, bulimia nervosa, and binge-eating disorder. The EDE also contains four subscales that measure dietary restraint, eating concern, weight concern, and shape concern. Using community and clinical samples, studies by Cooper, Cooper, and Fairburn (1989), Fairburn and Cooper (1993), Rizvi, Peterson, Crow, and Agras (1999), Williamson, Anderson, Jackman, and Jackson (1995), and Wilson and Smith (1989) reported internal consistency coefficients for the EDE scales ranging from .76 to .90, test-retest reliability correlations for diagnostic items ranging from .83 to .97, and interrater reliability kappas ranging from .83 to .99.

To assess the interrater reliability of the EDE in this study, a subset (25%) of EDE interviews were audiotaped and rated by a second interviewer who was unaware of the original diagnosis, resulting in acceptable interrater agreement (kappas ranged from .92 to 1.00). The EDE was administered to 346 of the 367 participants to ascertain eating-disorder diagnosis.

The SCID (Spitzer et al., 1990) served as the "gold standard" against which our self-report diagnostic scale was compared for a subset of participants (the 3 anorexia nervosa patients from the inpatient unit and 18 participants from the affect study) where we could not administer the more detailed EDE. The SCID is a standardized interview that assesses psychiatric status for major Axis I psychiatric disorders. Research has provided evidence of the reliability diagnoses of the SCID, with interrater reliability agreement kappas ranging from .70 to 1.00 and test-retest reliability of the eating-disorders sections ranging from .82 to .90 in community and clinical

samples (Pike, Loeb, & Walsh, 1995; Segal, Hersen, & Van Hasselt, 1994; Stukenberg, Dura, & Kiecolt-Glaser, 1990).

Responses to the EDE and SCID interviews were used to group participants into four diagnostic categories: *DSM-IV* anorexia nervosa ($n = 15$), *DSM-IV* bulimia nervosa ($n = 31$), *DSM-IV* binge-eating disorder ($n = 48$), and noneating disordered controls ($n = 273$).

Yale-Brown-Cornell Eating Disorder Scale. The Yale-Brown-Cornell Eating Disorder Scale (YBC-EDS; Mazure, Halmi, Sunday, Romano, & Einhorn, 1994) is an 84-item structured interview that assesses obsessions and rituals relating to food, eating, weight, and shape issues. Research has supported the reliability of the YBC-EDS measures, with internal consistency coefficients ranging from .82 to .90 and interrater agreement kappas ranging from .80 to 1.00 in clinical samples (Mazure et al., 1994; Sunday, Halmi, & Einhorn, 1995). All participants completed the YBC-EDS except the 38 participants from the affect study, 109 participants from the longitudinal risk factor study, and the 3 anorexic participants from the inpatient treatment unit.

Three-Factor Eating Questionnaire. The Three-Factor Eating Questionnaire (TFEQ; Stunkard & Messick, 1985) is a 54-item questionnaire with three subscales measuring cognitive (dietary) restraint, perceived hunger, and emotionally based disinhibition of eating. The reliability and validity of the TFEQ subscales has been supported by research indicating that the internal consistency coefficients ranged between .85 and .93, and that the subscales correlated with dietary intake and discriminated between dieters and nondieters in both community and clinical samples (French, Jeffery, & Wing, 1994; Laessle, Tuschl, Kotthaus, & Pirke, 1989; Stunkard & Messick, 1985). The 38 participants from the affect study, 109 participants from the longitudinal risk-factor study, and the 3 anorexic participants from the inpatient unit did not complete the TFEQ.

Eating Disorder Diagnostic Scale. Items assessing the *DSM-IV* diagnostic criteria for anorexia nervosa, bulimia nervosa, and binge-eating disorder were developed and revised following the procedures described in Study 1. Responses to the EDDS were used to group participants into four diagnostic categories: *DSM-IV* anorexia nervosa ($n = 18$), *DSM-IV* bulimia nervosa ($n = 39$), *DSM-IV* binge-eating disorder ($n = 57$), and noneating disordered ($n = 253$). As noted above, because some researchers may desire a broadband measure of eating pathology, an overall symptom composite was formed by standardizing all items (to control for the effects of the different response formats) and then summing across all items (except the height and birth control pill items). Thus, the symptom composite reflected each participant's overall level of eating pathology.

Procedures

At baseline, all participants completed the EDDS prior to partaking in the structured diagnostic interview (either the EDE or the SCID). This order was chosen because concepts such as "binge eating episodes" are defined in the interview, and we wanted to minimize the possibility that completing the interview would influence how participants responded to the EDDS. All interviews were conducted by clinical assessors with either a bachelor's, master's, or doctorate in psychology. Clinical assessors attended 16 hr of training, wherein structured interview skills were taught, diagnostic criteria for eating disorders were reviewed, simulated interviews were observed, and interviews were role played. Assessors had to demonstrate an interrater agreement ($\kappa > .80$) with experts with tape-recorded interviews before they were allowed to collect data. After the interview, clinical assessors measured height and weight. Participants from the first four eating-disorder studies then completed the TFEQ and the YBC-EDS. A randomly selected subset of participants ($N = 55$) from the longitudinal course of eating disorders study was asked to complete the EDDS 1 week after coming into the laboratory for their structured interview to provide data on the test-retest reliability of this scale. Participants were compensated for completing these measures (compensation ranged from \$15 to \$50, depending on the study).

Results

Observed means and standard deviations for all continuous measures used in Study 2 are reported in Table 1.

Test-Retest Reliability

The 1-week test-retest kappa coefficient was .95 for anorexia nervosa diagnoses, and the overall accuracy rate was .98. Kappa represents a chance-corrected level of agreement between two nominal variables, which, in this case, were the EDDS diagnoses at Time 1 and Time 2. Accuracy is the proportion of individuals in the sample that were labeled as disordered or nondisordered by the EDDS at both Time 1 and Time 2. The 1-week test-retest kappa coefficient was .71 for bulimia nervosa diagnoses, and the overall accuracy rate was .91. For binge-eating disorder, the 1-week test-retest kappa coefficient was .75, and the overall accuracy rate was .89.

We also examined the temporal stability of the EDDS overall symptom composite. The correlation coefficient reflecting the 1-week test-retest reliability was .87 for this composite.

Internal Consistency

The internal consistency of the EDDS symptom composite was assessed by calculating Cronbach's alpha for the standardized items that make up this score.¹ Cronbach's alpha for the symptom composite was .91 in the full sample and .86 for the subset of participants ($N = 55$) who completed the EDDS 1 week after coming into the lab.

Criterion Validity

The criterion validity of the EDDS was examined by testing whether, for each eating disorder, this scale accurately distinguished between interview-identified participants with the disorder and those without an eating disorder. Table 2 presents the kappa coefficient, sensitivity, specificity, positive predictive value, negative predictive value, and overall accuracy for each eating disorder. Kappa represents the chance-corrected level of agreement between two nominal variables, which, in this case, were the

interview diagnosis and the EDDS diagnosis for each eating disorder. Sensitivity reflects the proportion of individuals with a positive interview diagnosis who were correctly identified by the EDDS. Specificity reflects the proportion of individuals with a negative interview diagnosis who were correctly identified by the EDDS. Positive predictive value represents the proportion of individuals who were classified as having a positive diagnosis by the EDDS who actually met criteria for the diagnosis on the structured interview. Negative predictive value represents the proportion of individuals who were classified as having a negative diagnosis by the EDDS who actually did not meet criteria for the diagnosis on the structured interview. Accuracy is the proportion of individuals for whom the negative and positive EDDS diagnoses matched the actual interview diagnoses.

The kappa coefficient reflecting the agreement between the diagnoses from the structured interview and the EDDS was .93 for anorexia nervosa; and the sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were all above .93. The kappa coefficient that denoted diagnostic agreement between the structured interview and EDDS was .81 for bulimia nervosa; and the sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were all above .81. Finally, the kappa coefficient for diagnostic agreement between the interview and EDDS diagnoses was .74 for binge-eating disorder; and the sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were all above .77.

Convergent Validity

In an effort to generate evidence of convergent validity, we tested whether the EDDS-identified groups with eating disorders showed the expected elevations on validated measures of eating disturbances relative to EDDS-identified individuals with noneating disorders. Planned contrasts in analysis of variance (ANOVA) models that compared each eating-disorder group with the noneating-disorder group were used to accomplish this aim. We hypothesized that the three eating-disorder groups would show elevations on the validated measures of eating disturbances relative to the nondisorder group with two exceptions. First, because anorexia nervosa is marked by extreme caloric restriction, we did not expect this group to show elevations in disinhibited eating. Second, we did not expect individuals with binge-eating pathology to show elevations in dietary restraint relative to noneating-disorder controls because this disorder is characterized by excessive caloric intake in the absence of compensatory behaviors (such as extreme dieting).

The means and standard deviations for each of the eating disturbance scales across the various eating-disorder and control groups are presented in Table 3, along with the results of the planned contrasts and the percentage of variance accounted for in

Table 1
Means and Standard Deviations for All Continuous Measures Used in Study 2

Continuous measures	<i>M</i>	<i>SD</i>
Eating Disorder Examination		
Restraint	2.19	1.55
Eating concern	1.54	1.45
Weight concern	2.97	1.53
Shape concern	3.22	1.55
Yale-Brown-Cornell		
Eating and weight preoccupations	6.25	4.64
Eating and weight rituals	5.53	5.01
Three-Factor Eating Questionnaire		
Cognitive restraint	10.95	5.11
Hunger	7.76	3.98
Disinhibition	10.28	4.31
EDDS Symptom Composite	0.00	11.32

Note. EDDS = Eating Disorder Diagnostic Scale. $N = 367$.

¹ We considered conducting confirmatory factor analysis, but decided against this because the symptom overlap between eating disorders (e.g., the items on the importance of weight and shape on self-evaluation are used for diagnoses of anorexia nervosa and bulimia nervosa) would require cross-loadings that would cloud the interpretation of the factors. Moreover, the primary aim of this scale was to provide eating-disorder diagnoses, and we felt that a focus on the factorial structure would eclipse the main purpose of this scale.

Table 2
Agreement Between the EDDS Diagnoses and Structured Psychiatric Interview Diagnoses

Eating disorder	κ	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Accuracy
Anorexia nervosa	.93	.93	1.00	.93	1.00	.99
Bulimia nervosa	.81	.81	.98	.86	.97	.96
Binge-eating disorder	.74	.77	.96	.80	.95	.93

Note. EDDS = Eating Disorder Diagnostic Scale. $N = 367$.

each outcome by the grouping variable.² Statistical significance was assessed with the least significant difference test. As predicted, the eating-disorder groups generally showed elevations in dietary restraint, eating, weight and shape concerns, eating and weight preoccupations and rituals, cognitive restraint, hunger, and disinhibition on the validated measures of eating pathology. These effects accounted for between 6% to 31% of the variance in these outcomes. Consistent with expectations, the anorexia nervosa group did not show elevations on disinhibited eating, and the binge-eating disorder group did not show elevations on the two measures of dietary restraint. However, the anorexia nervosa group did not report heightened levels of hunger or weight and shape concerns, the bulimia nervosa group did not show elevations on the cognitive restraint scale, and the binge-eating disorder group did not evidence greater eating and weight rituals.

Finally, in an effort to provide evidence for the convergent validity of the EDDS symptom composite, we tested whether this composite was positively correlated with the validated measures of eating disturbances. As reflected in Table 4, the symptom composite showed significant positive correlations with the validation measures of dietary restraint, eating, weight and shape concerns, eating and weight preoccupations and rituals, hunger, and disinhibition. The one exception was that the EDDS symptom composite was not significantly correlated with the cognitive restraint scale. The significant relations accounted for between 13% and 43% of the variance in these validated measures of eating disturbances.

Discussion

The aim of this study was to develop a brief self-report scale for diagnosing anorexia nervosa, bulimia nervosa, and binge-eating disorder, and to provide evidence of its reliability and validity. We adapted items from validated structured interviews that assess eating pathology to create a one-page self-report scale. Collectively, findings from our two studies suggested that this new scale was both reliable and valid.

Evidence of the Reliability of the EDDS

Results suggested that the EDDS possessed satisfactory test-retest reliability in this investigation. The overall level of agreement for eating-disorder diagnoses between the two administrations of the EDDS separated by a period of 1 week was 98% for anorexia nervosa, 91% for bulimia nervosa, and 89% for binge-eating disorder, suggesting strong concordance between diagnoses generated by the EDDS over time. Test-retest kappa coefficients were good to excellent according to the criteria proposed by Fleiss (1981). These values reflect a reasonably high level of test-retest

reliability for a one-page instrument. Moreover, this scale compares favorably with validated psychiatric interviews such as the SCID, which had test-retest kappa coefficients ranging from .80 to .90 for eating-disorder diagnoses (Pike et al., 1995) and an average test-retest kappa coefficient across Axis I disorders of .46 in past clinical and community studies (e.g., Williams et al., 1992). We cannot compare the test-retest reliability of our new scale to that for the EDE because the test-retest reliability of EDE-generated diagnoses has not yet been reported. Nonetheless, it was noteworthy that the test-retest reliability coefficients for bulimia nervosa and binge-eating disorder were lower than the coefficient for anorexia nervosa. This pattern of findings likely resulted from the challenge of accurately measuring binge-eating frequency. Alternatively, there is some possibility that bulimia nervosa and binge-eating disorder are simply less temporally stable than anorexia nervosa.

Data also indicated that the EDDS symptom composite showed high test-retest reliability over a 1-week interval ($r = .87$). This estimate also compares favorably to those for other well-validated continuous measures of eating pathology, such as the test-retest coefficient of .95 for the Bulimia Test—Revised (Thelen et al., 1991), particularly given that the EDDS is much briefer.

The symptom composite also evidenced acceptable internal consistency across items (mean $\alpha = .89$). Therefore, in addition to using the EDDS as a brief measure of eating-disorder diagnoses, it appears that it may also be useful as a continuous measure of overall eating-disorder symptomatology.

Evidence of the Validity of the EDDS

The current findings also provided considerable evidence for the validity of the EDDS. First, the content validation study that used expert raters suggested that the relevant *DSM-IV* diagnostic criteria for the three eating disorders were included on the EDDS and that no irrelevant information was assessed.

Second, agreement between the eating-disorder diagnoses from the EDDS and those from the structured interviews was 99% for anorexia nervosa, 96% for bulimia nervosa, and 93% for binge-eating disorder, which represented good to excellent concordance. Collectively, these results suggest that the EDDS possessed adequate criterion validity in this investigation. The agreement be-

² The sample size ($N = 217$) for the convergent validity analyses was smaller because the 3 anorexic patients recruited from the inpatient psychiatric treatment unit, the 38 participants recruited from the affect study, and the 109 participants recruited from the longitudinal risk factor study did not complete the EDE subscales, YBC-EDS, or TFEQ.

Table 3
Mean Differences Between EDDS-Identified Eating-Disorder Groups and Controls on Validated Eating Pathology Measures

Eating pathology measures	Anorexia nervosa (<i>n</i> = 18)		Bulimia nervosa (<i>n</i> = 39)		Binge-eating disorder (<i>n</i> = 57)		Non-eating disordered controls (<i>n</i> = 103)		% variance explained
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Eating Disorder Examination									
Restraint	3.24	1.78 _a	2.98	1.37 _a	2.01	1.15	1.84	1.60 _b	11
Eating concern	2.20	1.43 _a	2.36	1.55 _a	1.83	1.29 _a	0.99	1.28 _b	14
Weight concern	2.43	1.93	3.63	1.27 _a	3.60	1.16 _a	2.48	1.53 _b	11
Shape concern	2.83	1.83	4.13	1.28 _a	3.95	1.06 _a	2.56	1.50 _b	18
Yale-Brown-Cornell									
Eating and weight preoccupations	9.27	4.75 _a	8.95	3.07 _a	6.36	4.68 _a	4.77	4.46 _b	15
Eating and weight rituals	9.46	4.93 _a	7.82	4.74 _a	4.83	4.64	4.40	4.75 _b	12
Three-Factor Eating Questionnaire									
Cognitive restraint	14.13	3.39 _a	12.08	4.24	9.60	5.13	10.62	5.38 _b	6
Hunger	6.27	3.73	9.95	3.88 _a	9.60	3.38 _a	6.40	3.65 _b	18
Disinhibition	5.60	3.96	12.79	2.52 _a	13.14	2.20 _a	8.85	4.29 _b	31

Note. EDDS = Eating Disorder Diagnostic Scale. Means with different subscripts are statistically significantly different according to the least significant difference test. *N* = 217.

tween the diagnoses generated by the EDDS and the "gold standard" approached the upper limit of what could be expected given the reliability of the measures. Moreover, the observed validity coefficients compare favorably to those for the SCID (Kranzler, Kadden, Babor, & Tennen, 1996). Nonetheless, it was noteworthy that the concordance between the diagnoses from the EDDS and the interview was somewhat lower for binge-eating disorder than it was for anorexia nervosa and bulimia nervosa. There are at least two factors that might have contributed to this finding. First, the 6-month duration criterion for binge-eating disorder symptoms is longer than the duration criteria for bulimia nervosa or anorexia nervosa, which may make it more difficult for people to recall their symptom duration accurately (Henry, Moffitt, Caspi, Langley, & Silva, 1994). Second, more qualifiers are required for binge episodes to meet *DSM-IV* diagnostic criteria for binge-eating disorder than is the case for bulimia nervosa. In addition to the binge-eating requirements for a diagnosis of bulimia nervosa (the participant

must consume a large amount of food and feel out of control), binge-eating disorder requires that at least three additional features are present during binge eating (e.g., eating more rapidly than normal, eating alone because of embarrassment about eating behaviors) and that binge eating result in marked distress. The reduced concordance for binge-eating disorder may be due to the fact that participants had to report the exact same responses to more questions on the scale and in the interview.

Third, data also suggested that the EDDS possessed convergent validity in this investigation, in that those identified as having an eating disorder based on the EDDS generally showed elevated scores on validated measures of eating disturbances relative to EDDS-identified with no eating disorders. As hypothesized, EDDS-identified eating-disorder groups reported more dietary restraint, hunger, and disinhibited eating, as well as greater eating, weight and shape concerns and rituals, than did EDDS-identified controls. Moreover, the EDDS symptom composite correlated positively with these same validated measures of eating pathology. The magnitude of the significant relations ranged from medium to large effect sizes according to the criteria proposed by Cohen (1988). There were a few nonsignificant group differences that were predicted a priori. Consistent with expectations, the binge-eating disorder group did not evidence heightened dietary restraint. Again, individuals with binge-eating disorder were not expected to show elevations in dietary restraint because this disorder is characterized by excessive caloric intake in the absence of compensatory behaviors. Also as predicted, the anorexic group did not show elevated disinhibition relative to controls. Presumably, this group did not show greater disinhibited eating because anorexia nervosa is marked by extreme caloric restriction.

However, there were also three unexpected findings in the convergent validity analyses. First, the anorexic group did not report greater hunger or elevated weight and shape concerns compared with non-eating disorder controls. This may reflect the extreme denial that characterizes anorexia nervosa (Vitousek, Watson, & Wilson, 1998). Indeed, the fact that EDDS-identified anorexic participants denied hunger when they had a mean BMI

Table 4
Pearson Product Correlations Between the EDDS Symptom Composite and the Validated Eating Pathology Measures

Eating pathology measures	Correlation with EDDS symptom composite
Eating Disorder Examination	
Restraint	.36*
Eating concern	.54*
Weight concern	.57*
Shape concern	.66*
Yale-Brown-Cornell	
Eating and weight preoccupations	.47*
Eating and weight rituals	.36*
Three-Factor Eating Questionnaire	
Cognitive restraint	.10
Hunger	.53*
Disinhibition	.63*

Note. EDDS = Eating Disorder Diagnostic Scale, *N* = 217.
* *p* < .001.

score of 15.5 (mean weight = 87 lbs. and mean height = 5 ft 3 in.) clearly suggests they were in denial. Similarly, it seems unlikely that these individuals would have restricted their caloric intake severely enough to have achieved this low weight unless they had elevated weight and shape concerns. These findings underscore the difficulty of measuring a disorder characterized by denial with a self-report instrument. Second, EDDS-identified binge-eating disorder individuals did not report heightened eating and weight rituals relative to controls on the YBC subscale. Perhaps individuals with binge-eating disorder do not engage in elaborate eating and weight rituals (e.g., constant weighing and compulsive calorie counting) because this eating disorder is less rooted in the pursuit of the thin ideal than is anorexia and bulimia nervosa. Third, the bulimic group did not report greater dietary restraint than controls on one of the two restraint measures, and this same scale did not correlate significantly with the EDDS symptom composite. One possibility suggested by the fact that the effects were larger for the EDE restraint scale than for the TFEQ cognitive restraint measure was that the former is simply more sensitive. Alternatively, it has been proposed that the TFEQ taps successful dieting (Heatherton, Herman, Polivy, King, & McGree, 1988), and it may be that individuals with eating disorders, such as bulimic and binge-eating disorder individuals, do not engage in effective dieting. Finally, the nonsignificant finding for the TFEQ restraint scale may also have been related to the challenge of measuring dietary restriction within the context of regular binge eating.

Limitations

Although this study utilized a large multisite sample, incorporated structured psychiatric interviews, and attempted to provide a wide range of evidence for the reliability and validity of the EDDS, several limitations deserve comment. First, the fact that only female individuals were included in the present sample obviously precludes generalization to male individuals. Future research will be necessary to establish the reliability and validity of the EDDS for male individuals. Second, the fact that participants for this investigation were drawn from a variety of sources makes it somewhat difficult to know to which population these findings can be generalized. Thus, caution should be exercised in generalizing these results. Third, it would have been ideal to have replicated these psychometric analyses in an independent sample to provide even greater confidence in the stability of these reliability and validity estimates. Fourth, the test-retest coefficients for diagnoses should be interpreted with caution because of the moderate sample size for those analyses. On a related note, the number of participants in some of the eating-disorder groups was modest (e.g., there were only 15 participants with anorexia nervosa), which limits the confidence that can be placed in the findings.

Conclusion

This study provided evidence that the EDDS showed content validity; that the EDDS-derived eating-disorder diagnoses possessed test-retest reliability, criterion validity, and convergent validity; and that the EDDS symptom composite showed test-retest reliability, internal consistency, and convergent validity. Collectively, results from this preliminary investigation indicated that the EDDS showed reasonable psychometric qualities. The benefits of this scale are that it can be completed quickly and

easily, and that it is inexpensive because it is not necessary to train or pay interviewers. Thus, it appears that this scale might ultimately prove useful for the assessment of eating disorders in etiologic, prevention, and treatment research applications, as well as in traditional clinical settings, where structured psychiatric interviews are less feasible.

References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Brody, M. J., Walsh, B. T., & Devlin, M. (1994). Binge eating disorder: Reliability and validity of a new diagnostic category. *Journal of Consulting and Clinical Psychology, 62*, 381-386.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cooper, Z., Cooper, P. J., & Fairburn, C. G. (1989). The validity of the eating disorder examination and its subscales. *British Journal of Psychiatry, 154*, 807-812.
- Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders, 16*, 363-370.
- Fairburn, C. G., & Cooper, Z. (1993). The eating disorder examination (12th ed.). In C. Fairburn & G. Wilson (Eds.), *Binge eating: Nature, assessment, and treatment* (pp. 317-360). New York: Guilford Press.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.
- French, S. A., Jeffery, R. W., & Wing, R. R. (1994). Food intake and physical activity: A comparison of three measures of dieting. *Addictive Behaviors, 19*, 401-409.
- Garfinkel, P. E., Lin, E., Goering, P., Spegg, C., Goldbloom, D. S., Kennedy, S., Kaplan, A. S., & Woodside, D. B. (1995). Bulimia nervosa in a Canadian community sample: Prevalence and comparison of subgroups. *American Journal of Psychiatry, 152*, 1052-1058.
- Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*, 238-247.
- Heatherton, T. F., Herman, C. P., Polivy, J., King, G. A., & McGree, S. T. (1988). The (mis)measurement of restraint: An analysis of conceptual and psychometric issues. *Journal of Abnormal Psychology, 97*, 19-28.
- Henry, B., Moffitt, T. E., Caspi, A., Langley, J., & Silva, P. A. (1994). On the "remembrance of things past": A longitudinal evaluation of the retrospective method. *Psychological Assessment, 6*, 92-101.
- Keller, M. B., Herzog, D. B., Lavori, P. W., Bradburn, I. S., & Mahoney, E. M. (1992). The natural history of bulimia nervosa: Extraordinarily high rates of chronicity, relapse, recurrence, and psychosocial morbidity. *International Journal of Eating Disorders, 12*, 1-9.
- Kranzler, H. R., Kadden, R., Babor, T., & Tennen, H. (1996). Validity of the SCID in substance abuse patients. *Addiction, 91*, 859-868.
- Laessle, R. G., Tuschl, R. J., Kotthaus, B. C., & Pirke, K. M. (1989). A comparison of the validity of three scales for the assessment of dietary restraint. *Journal of Abnormal Psychology, 98*, 504-507.
- Mazure, C. M., Halmi, K. A., Sunday, S. R., Romano, S. J., & Einhorn, A. M. (1994). Yale-Brown-Cornell Eating Disorder Scale: Development, use reliability, and validity. *Journal of Psychiatric Research, 28*, 425-445.
- Mintz, L. B., O'Halloran, M. S., Mulholland, A. M., & Schneider, P. A. (1997). Questionnaire for eating disorder diagnoses: Reliability and validity of operationalizing DSM-IV criteria into a self-report format. *Journal of Counseling Psychology, 44*, 63-79.
- Newman, D. L., Moffitt, T. E., Caspi, A., Magdol, L., Silva, P. A., & Stanton, W. R. (1996). Psychiatric disorder in a birth cohort of young adults: Prevalence, comorbidity, clinical significance, and new case incidence from ages 11-21. *Journal of Consulting and Clinical Psychology, 64*, 552-562.

- Pike, K. M., Loeb, K., & Walsh, T. (1995). Binge eating and purging. In D. Allison (Ed.), *Handbook of assessment methods for eating behaviors and weight-related problems* (pp. 303-346). Thousand Oaks, CA: Sage.
- Rizvi, S. L., Peterson, C. B., Crow, S. J., & Agras, W. S. (1999). Test-retest reliability of the eating disorder examination. *International Journal of Eating Disorders*, 26, 406-413.
- Segal, D. L., Hersen, M., & Van Hasselt, V. B. (1994). Reliability of the Structured Clinical Interview for *DSM-III-R*: An evaluative review. *Comprehensive Psychiatry*, 35, 316-327.
- Smolak, L., Levine, M. P., & Striegel-Moore, R. (1996). *The developmental psychopathology of eating disorders*. Mahwah, NJ: Erlbaum.
- Spitzer, R. L., Devlin, M., Walsh, B. T., Hasin, D., Wing, R., Marcus, M., Stunkard, A., Wadden, T., Yanovski, S., Agras, W. S., Mitchell, J., & Nonas, C. (1992). Binge eating disorder: A multisite field trial of the diagnostic criteria. *International Journal of Eating Disorders*, 11, 191-203.
- Spitzer, R. L., Williams, J. B., Gibbon, M., & First, M. B. (1990). *Structured Clinical Interview for DSM-III-R (SCID)*. Washington, DC: American Psychiatric Press.
- Spitzer, R. L., Yanovski, S., Wadden, T., Wing, R., Marcus, M. D., Stunkard, A., Devlin, M., Mitchell, J., Hasin, D., & Horne, R. (1993). Binge eating disorder: Its further validation in a multisite study. *International Journal of Eating Disorders*, 13, 137-153.
- Stukenberg, K. W., Dura, J. R., & Kiecolt-Glaser, J. K. (1990). Depression Screening Scale validation in an elderly, community-dwelling population. *Psychological Assessment*, 2, 134-138.
- Stunkard, A. J., & Messick, S. (1985). The Three Factor Eating Questionnaire to measure dietary restraint, disinhibition, and hunger. *Journal of Psychosomatic Research*, 29, 71-83.
- Sunday, S. R., Halmi, K. A., & Einhorn, A. (1995). The Yale-Brown-Cornell Eating Disorder Scale: A new scale to assess eating disorder symptomatology. *International Journal of Eating Disorders*, 18, 237-245.
- Telch, C., & Stice, E. (1998). Psychiatric comorbidity in a non-clinical sample of women with binge eating disorder. *Journal of Consulting and Clinical Psychology*, 66, 768-776.
- Thelen, M., Farmer, J., Wonderlich, S., & Smith, M. (1991). A revision of the Bulimia Test: The BULIT-R. *Psychological Assessment*, 3, 119-124.
- Vitousek, K., Watson, S., & Wilson, G. T. (1998). Enhancing motivation for change in treatment-resistant eating disorders. *Clinical Psychology Review*, 18, 391-420.
- Whitaker, A., Johnson, J., Shaffer, D., Rapoport, J. L., Kalikow, K., Walsh, B. T., Davies, M., Braiman, S., & Dolinsky, A. (1990). Uncommon troubles in young people: Prevalence estimates of selected psychiatric disorders in a nonreferred adolescent population. *Archives of General Psychiatry*, 47, 487-496.
- Williams, J. B., Gibbon, M., First, M., Spitzer, R. L., Davies, M., Borus, J., Howes, M., Kane, J., Pope, H., Rounsaville, B., & Wittchen, H. (1992). The Structured Clinical Interview for *DSM-III-R* (SCID): Multisite test-retest reliability. *Archives of General Psychiatry*, 49, 630-636.
- Williamson, D. A., Anderson, D. A., Jackman, L. P., & Jackson, S. R. (1995). Assessment of eating disordered thoughts, feelings, and behaviors. In D. B. Allison (Ed.), *Handbook of assessment methods for eating behaviors and weight-related problems: Measures, theory, and research* (pp. 347-386). Thousand Oaks, CA: Sage.
- Wilson, G. T., Heffernan, K., & Black, C. M. (1996). Eating disorders. In E. J. Mash & R. A. Barkley (Eds.), *Child psychopathology* (pp. 541-571). New York: Guilford Press.
- Wilson, G. T., & Smith, D. (1989). Assessment of bulimia nervosa: An evaluation of the eating disorders examination. *International Journal of Eating Disorders*, 8, 173-179.

Appendix A

Scoring Algorithm for Eating Disorder Diagnosis Scale (EDDS)

The scoring algorithm for EDDS eating disorder diagnoses parallels that used for the Eating Disorder Examination (EDE). Computerized scoring statements are ordered such that bulimia nervosa diagnoses preempt binge-eating disorder diagnoses, and anorexia nervosa diagnoses preempt bulimia nervosa diagnoses. The computer code is available from Eric Stice.

Anorexia Nervosa

A diagnosis of *DSM-IV* anorexia nervosa is made if an individual reports (a) height and weight data on EDDS Items 19 and 20 that result in a body mass index ($BMI = Kg/M^2$) of less than 17.5, (b) a fear of weight gain or becoming fat as indexed by a score of 4 or greater on EDDS Item 2, (c) undue influence of body weight or shape on self-evaluation as indexed by a score of 4 or greater on either EDDS Item 3 or 4, and (d) amenorrhea in postmenarcheal females as indexed by a 3 on EDDS Item 21. Following the EDE scoring algorithm, if an individual meets the first and fourth criteria above, it is not necessary for the individual to endorse the second and third criteria. Further, because oral contraceptives can result in a regular menstrual cycle, to be on the conservative side, participants who were taking oral contraceptives that met the low weight criteria were coded as amenorrheic. This approach is also used in the EDE.

Bulimia Nervosa

A diagnosis of *DSM-IV* bulimia nervosa is made if an individual reports (a) regular eating binges marked by a perceived loss of control and the consumption of a large amount of food as indexed by a response of yes to EDDS Item 5, a yes to EDDS Item 6, and a response of greater than 2 on EDDS Item 8; (b) regular use of compensatory behaviors as indexed by a response of 8 or greater on the sum of EDDS Items 15, 16, 17, and 18; and (c) undue influence of body weight or shape on self-evaluation as indexed by a score of 4 or greater on either EDDS Item 3 or 4.

Binge-Eating Disorder

A diagnosis of *DSM-IV* binge-eating disorder is made if an individual reports (a) regular eating binges marked by a perceived loss of control and the consumption of a large amount of food as indexed by a response of yes to EDDS Item 5, a yes to EDDS Item 6, and a response of greater than 2 on EDDS Item 7; (b) an endorsement of at least three of the features that may be associated with binge eating as indexed by a yes response to at least three of the features described in EDDS Items 9, 10, 11, 12, and 13; (c) marked distress regarding binge eating as indexed by a yes response to EDDS Item 14; and (d) the absence of any compensatory behaviors as reflected by a 0 response to EDDS Items 15, 16, 17, and 18.

Appendix B

Eating Screen

Please carefully complete all questions.

Over the past 3 months . . .	Not at all		Slightly		Moderately		Extremely
1. Have you felt fat?	0	1	2	3	4	5	6
2. Have you had a definite fear that you might gain weight or become fat?	0	1	2	3	4	5	6
3. Has your weight influenced how you think about (judge) yourself as a person?	0	1	2	3	4	5	6
4. Has your shape influenced how you think about (judge) yourself as a person?	0	1	2	3	4	5	6
5. During the past 6 months have there been times when you felt you have eaten what other people would regard as an unusually large amount of food (e.g., a quart of ice cream) given the circumstances? YES NO							
6. During the times when you ate an unusually large amount of food, did you experience a loss of control (feel you couldn't stop eating or control what or how much you were eating)? YES NO							
7. How many DAYS per week on average over the past 6 MONTHS have you eaten an unusually large amount of food and experienced a loss of control? 0 1 2 3 4 5 6 7							
8. How many TIMES per week on average over the past 3 MONTHS have you eaten an unusually large amount of food and experienced a loss of control? 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14							

During these episodes of overeating and loss of control did you . . .

9. Eat much more rapidly than normal? YES NO
10. Eat until you felt uncomfortably full? YES NO
11. Eat large amounts of food when you didn't feel physically hungry? YES NO
12. Eat alone because you were embarrassed by how much you were eating? YES NO
13. Feel disgusted with yourself, depressed, or very guilty after overeating? YES NO
14. Feel very upset about your uncontrollable overeating or resulting weight gain? YES NO
15. How many times per week on average over the past 3 months have you made yourself vomit to prevent weight gain or counteract the effects of eating? 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
16. How many times per week on average over the past 3 months have you used laxatives or diuretics to prevent weight gain or counteract the effects of eating? 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
17. How many times per week on average over the past 3 months have you fasted (skipped at least 2 meals in a row) to prevent weight gain or counteract the effects of eating? 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
18. How many times per week on average over the past 3 months have you engaged in excessive exercise specifically to counteract the effects of overeating episodes? 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
19. How much do you weigh? If uncertain, please give your best estimate. ____lb
20. How tall are you? __ft __in.
21. Over the past 3 months, how many menstrual periods have you missed? 1 2 3 4 na
22. Have you been taking birth control pills during the past 3 months? YES NO