Questionnaire for Eating Disorder Diagnoses: Reliability and Validity of Operationalizing *DSM-IV* Criteria Into a Self-Report Format

Laurie B. Mintz
University of Missouri—Columbia

M. Sean O'Halloran
University of Northern Colorado

Amy M. Mulholland and Paxton A. Schneider
University of Missouri—Columbia

The Questionnaire for Eating Disorder Diagnoses (Q-EDD) operationalizes eating disorder criteria of the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders* and differentiates (a) between those with and without an eating disorder diagnosis, (b) among eating-disordered, symptomatic, and asymptomatic individuals, and (c) between those with anorexia and bulimia diagnoses. Three studies examined the Q-EDD's psychometric properties. Convergent validity was supported by correspondence between Q-EDD diagnoses and established inventory scores. Criterion validity was supported by high correspondence between Q-EDD and interview or clinician diagnoses. Incremental validity was supported by greater accuracy of Q-EDD diagnoses than those yielded by an established inventory. Test–retest reliability and interscorer agreement were very good. Future use is discussed.

Studies concerning the prevalence, correlates, and epidemiology of eating disorders among women have proliferated in recent years, among both counseling psychologists and others. Unfortunately, there are important shortcomings in the ways in which eating disorders are operationalized in this literature, particularly when nonclinical samples such as college students and community women are examined. In these samples, eating-disordered women are selected for

study by the use of pencil-and-paper self-report measures. There are two types of such self-report measures: (a) pre-existing inventories, such as the Eating Attitudes Test (EAT; Garner & Garfinkel, 1979) and the revised Bulimia Test (BULIT-R; Thelen, Farmer, Wonderlich, & Smith, 1991), and (b) questionnaires that are designed de novo for specific studies to operationalize American Psychiatric Association *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*) criteria (Patton & King, 1991). Both types are limited in several ways. Two problems with preexisting inventories are that (a) they are all based on outdated *DSM* criteria (e.g., the third edition [*DSM-III*; American Psychiatric Association, 1980] or the revised third edition [*DSM-III-R*; American Psychiatric Association, 1987] and (b) none can capture a range of disorders or provide differential diagnoses of eating disorders (Williamson, 1990; Williamson, Anderson, Jackman, & Jackson, 1995). In addition, none make differentiations within the group of individuals scoring in the nonpathological range on the inventory (i.e., between those with no symptoms and those with some symptoms of the disorder being diagnosed). Many problems also plague the use of de novo questionnaires, including a lack of generalizability across studies and the fact that they often only partially operationalize *DSM* criteria (Fairburn, Phil, & Beglin, 1990). Most important, very few researchers using either type of pencil-and-paper measure have confirmed such self-report diagnoses with clinical interviews. This is especially problematic when de novo questionnaires are used; there is little knowledge of the level of agreement between diagnoses yielded by such operationalized *DSM* questionnaires and expert ratings (e.g., clinician diagnoses; Fairburn et al., 1990). All of these interrelated issues are discussed subsequently. The goal of this research was to begin to overcome these problems by operationalizing the eating disorder criteria of the fourth edition of the *DSM* (*DSM-IV*; American Psychiatric Association, 1994) into a

self-report format and examining the reliability and validity obtained when using this questionnaire to make diagnostic differentiations that no current inventory is capable of making.

The *DSM–IV* lists three types of eating disorders: bulimia nervosa (bulimia), anorexia nervosa (anorexia), and eating disorders not otherwise specified (EDNOS). A summary of the *DSM–IV* criteria for anorexia and bulimia can be found in Table 1. The EDNOS diagnosis is given for "disorders of eating that do not meet the criteria for any specific Eating Disorder" (American Psychiatric Association, 1994, p. 550). The *DSM–IV* lists six examples of EDNOS; these examples and the names by which we refer to them are listed in the note to Table 2. Many of the eating disorders falling in this category can be thought of as atypical (i.e., a feature absent) or subthreshold (i.e., all features present but not at sufficient frequency) forms of anorexia and bulimia (e.g., menstruating anorexia is an atypical form of anorexia, whereas subthreshold bulimia is a subthreshold form of bulimia; Fairburn & Garner, 1986).

The two types of self-report instruments used in eating

disorder research (preexisting and de novo) differ from one another in important ways. Most preexisting inventories yield numerical scores and use cutoffs (i.e., scores above or below a certain number) for arriving at diagnoses. On the other hand, most questionnaires designed de novo for studies contain questions that operationalize *DSM* criteria, and, like the *DSM* itself, most use dichotomous decision rules (e.g., meeting or not meeting criteria) for arriving at diagnoses. In other words, these types of questionnaires do not yield continuous data (i.e., numerical scores) but instead yield nominal data (i.e., diagnostic categories) on the basis of *DSM*-guided dichotomous decision rules. As discussed later, rarely is reliability or validity data collected on operationalized questionnaires, whereas such information is readily available for preexisting inventories.

Many preexisting instruments are available that measure eating-disordered thoughts, feelings, and behaviors, including those with the specific aim of diagnosing eating disorders such as anorexia and bulimia (for complete reviews, see Pike, Loeb, & Walsh, 1995; Williamson et al., 1995). Perhaps the most widely used preexisting inventories are the

Table 1
Selected Questionnaire for Eating Disorder Diagnoses (Q-EDD) Items and Scoring Manual Decision Rules

DSM–IV Criterion	Description of Q-EDD items	Decision rule
Anorexia nervosa		
Body weight <85% expected	Height in feet–inches; weight in pounds	Calculations indicating BMI of 17.5 or below
Fear of gaining weight or becoming fat, although underweight	Two 5-point Likert items asking, "How afraid are you of becoming fat/gaining weight?"	BMI of 20 or below and 4 or 5 response to one of the questions
Missed three consecutive menstrual cycles	Yes–no question asking whether three menstrual cycles have been missed, not including pregnancy	Yes response
Bulimia nervosa		
Recurrent binge eating characterized by:		
Eating, in a discrete period of time, an amount of food definitely larger than most would eat under similar time frame and circumstances	Yes–no question exactly corresponding to *DSM–IV*: "Do you experience binge eating, meaning...?"	Yes response
Sense of lack of control during binge episodes	Yes–no question exactly corresponding to *DSM–IV*: "Do you have a sense of...?"	Yes response
Recurrent inappropriate compensatory behaviors, such as self-induced vomiting, fasting, excessive exercise	Yes–no questions asking, "Do you...?" for several behaviors, with the exception of excessive exercise, which was operationalized by yes–no questions concerning exercise interfering with important activities and exercising despite injury–medical complications	Yes for at least one of defined behaviors
Self-evaluation unduly influenced by weight and shape	5-point Likert item asking, "Does your weight and/or body shape influence how you feel about yourself?"	4 or 5 response to this item

Note. Selected *DSM–IV* criteria are summarized and paraphrased here. In addition, for anorexia, *DSM–IV* criteria specify that one must exhibit disturbance in the way one's body weight or shape is experienced, undue influence of weight or shape on self-evaluation, or denial of the seriousness of current low body weight. For bulimia, criteria specify that binge eating and inappropriate compensatory behaviors both occur, on average, at least twice a week for 3 months and that the disturbance does not occur exclusively during periods of anorexia nervosa. The *DSM–IV* specifies subtypes of anorexia (binge eating/purging type and restricting type) and bulimia (purging and nonpurging type). BMI requirements and questions operationalizing excessive exercise were taken from the *DSM–IV* narratives. A complete copy of the Q-EDD and the scoring manual are available from Laurie B. Mintz. *DSM–IV* = *Diagnostic and Statistical Manual of Mental Disorders,* 4th edition; BMI = body mass index.

Table 2
EDNOS Group Operational Definitions

EDNOS group[a]	Q-EDD operationalization
Menstruating anorexia	Meets all criteria for anorexia except that concerned with menstruation (i.e., answers no to question concerning missing three consecutive menstrual cycles)
Subthreshold bulimia[a]	Meets all criteria for bulimia, except either binge eating or inappropriate compensatory behavior does not meet frequency criteria
Nonbinging bulimia[b]	Normal-weight individual who does not binge eat but who engages in inappropriate compensatory behaviors at frequency specified in bulimia criteria (i.e., at least twice a week for last 3 months)
Binge-eating disorder[c]	Binge eats at least twice a week for last 6 months, feels out of control during binge, engages in no inappropriate compensatory behaviors, does not diet, does not take appetite control pills

Note. The *DSM–IV* lists six examples of EDNOS on page 550: (a) "All the criteria for Anorexia Nervosa are met except that the individual has regular menses" (menstruating anorexia); (b) "all of the criteria for Anorexia Nervosa are met except that, despite significant weight loss, the individual's current weight is in the normal range" (normal weight anorexia); (c) "all of the criteria for Bulimia Nervosa are met except that the binge eating and inappropriate compensatory mechanisms occur at a frequency of less than twice a week or for a duration of less than three months" (subthreshold bulimia); (d) "the *regular* use of inappropriate compensatory behavior by an individual of normal body weight after eating small amounts of food" (nonbinging bulimia); (e) "repeatedly chewing and spitting out, but not swallowing, large amounts of food" (chew-spitting); and (f) "recurrent episodes of binge eating in the absence of the *regular* use of inappropriate compensatory behaviors characteristic of Bulimia Nervosa" (binge-eating disorder). The term *regular* is not emphasized in the *DSM–IV;* emphasis was added here to make the point that what was meant by *regular* needed to be operationalized by Q-EDD decision rules. Also, with the exception of binge-eating disorder, the labels in parentheses are not included in the *DSM–IV* but were coined by us. EDNOS = eating disorders not otherwise specified; Q-EDD = Questionnaire for Eating Disorder Diagnoses; *DSM– IV = Diagnostic and Statistical Manual of Mental Disorders,* 4th edition.
[a] Because of the response ranges on the Q-EDD, this would be a person who engages in inappropriate compensatory behaviors between once per week and once per month or binges less than twice per week or for less than 3 months. [b] For nonbinging bulimia, inappropriate compensatory behaviors were the same as those listed for bulimia (vomiting, laxatives, enema, and fasting) with the exception of excessive exercise; preliminary interviews indicated that women who met only the criteria for excessive exercise were athletes. Nonbinging bulimics have been called "purgers" in past research, but the current name was chosen because of the inclusion of fasting as an inappropriate compensatory behavior. [c] Because binge-eating disorder is included in Appendix B of the *DSM–IV* as a "Criteria Set Provided for Further Study," and because of the existence of an instrument that fully operationalizes *DSM–IV* Appendix B binge-eating disorder criteria (Spitzer et al., 1992), only the criteria concerning binge-eating frequency, lack of control, and lack of inappropriate compensatory behaviors were assessed by the Q-EDD because of their high effect on prevalence findings (see Spitzer et al., 1993, for a full discussion). For the sake of compatibility with the question assessing the bulimia criteria regarding frequency of binges (i.e., so that two separate questions on binge-eating frequency were not asked), frequency of binges was assessed by times per week rather than the episodes per week discussed in Appendix B of the *DSM–IV.* Although there is debate concerning the overlap among subthreshold bulimia and binge-eating disorder, the use of dieting by women with binge-eating disorder, and the taking of appetite control pills by women with binge-eating disorder, strict definitions were adhered to for Q-EDD purposes.

EAT (Garner & Garfinkel, 1979), the BULIT-R (Thelen et al., 1991), the Eating Disorder Inventory (EDI; Garner, Olmstead, & Polivy, 1983), and, more recently, the EDI-2 (Garner, 1991). The EAT was originally designed as a measure of the symptoms of anorexia defined by Feigner and colleagues (1972). (Garner & Garfinkel, 1979). Because of this, EAT items no longer reflect anorexia as defined in the *DSM–IV;* many of the scale's items relate to bulimia, with binge eating, self-induced vomiting, and lax-

ative abuse all being covered (Garner, Olmstead, Bohr, & Garfinkel, 1982). When used in nonclinical samples, the EAT has been found to have a high false-positive rate for capturing anorexia (i.e., many of those diagnosed by the EAT as anorexic are not found to be so on interview; Johnsone-Sabine, Wood, & Patton, 1988; Meadows, Palmer, Newball, & Kenrick, 1986). High EAT scorers in nonclinical samples have been found to include a range of nonanorexic individuals such as bulimics, subclinical an-

orexics, purgers, and obsessional dieters (Button & White-house, 1981; Clarke & Palmer, 1983; King, 1986). The EAT's authors have thus warned against using it for diagnosing anorexia in nonclinical samples and have stated that high EAT scores in such samples may instead indicate the presence of "disturbed eating patterns" (Garner et al., 1982, p. 877). Nevertheless, perhaps because it is the only preexisting inventory associated with diagnosing anorexia, the EAT is still widely used as a measure of anorexia in research conducted with nonclinical samples.

The BULIT-R is another widely used preexisting inventory. Although it was designed with *DSM–III–R* criteria for bulimia, recent research demonstrates that the BULIT-R is able to differentiate *DSM–IV*-defined bulimics from non-bulimics (Thelen, Mintz, & Vander Wal, 1996). However, the BULIT-R cannot make distinctions within the non-bulimic group. In other words, the nonbulimic group would include individuals with anorexia and EDNOS, as well as those with no eating disorders, none of whom would be differentiated from one another (Thelen et al., 1996). Hence, the BULIT-R is useful only for research on bulimic individuals.

The EDI and EDI-2 measure eating disorder symptomatology, and elevated scores on two subscales (i.e., Drive for Thinness and Bulimia) indicate a high likelihood of engaging in eating-disordered behaviors (Garner, 1991). Nevertheless, no specific information is provided on which behaviors are engaged in, and the subscale from which a bulimia diagnosis is sometimes made contains only questions about self-induced vomiting, even though a range of inappropriate compensatory behaviors constitutes the *DSM* criteria for bulimia (e.g., fasting and laxative abuse).

There thus remain major gaps in the ability to screen for eating disorders in nonclinical samples when using preexisting inventories. Specifically, there is no *DSM–IV*-based instrument that captures the diagnoses of anorexia or EDNOS, nor is there a *DSM–IV*-based instrument that identifies eating-disordered individuals in general (i.e., the combined group of all individuals with a *DSM–IV* eating disorder diagnosis). Thus, no inventory can differentially diagnose individuals with *DSM–IV* eating disorders from those without such disorders. Likewise, none of the existing instruments are able to differentially diagnose bulimia and anorexia (Williamson, 1990; Williamson et al., 1995).

Another potential problem is that none of these inventories make distinctions within the non-eating-disordered group. For example, with the EAT or the BULIT-R, an individual falling below the diagnostic cutoff score by 1 point (i.e., indicating endorsement of many eating-disordered behaviors) would be considered the same as an individual receiving the lowest possible score on the inventory (i.e., indicating endorsement of no eating-disordered behaviors). Indeed, both individuals would be considered as not having eating disorders and hence contrasted with eating-disordered individuals in studies of correlates and epidemiology. In sum, no distinctions can be made within the group of individuals scoring in the non-eating-disordered range on preexisting inventories.

In contrast, a few researchers using operationalized *DSM*

questionnaires (e.g., Hesse-Biber, 1992; Kurtzman, Yager, Landsverk, Wiesmeier, & Bodurka, 1989; Schotte & Stunkard, 1987; Striegel-Moore, Silberstein, Frensch, & Rodin, 1989; Zuckerman, Colby, Ware, & Lazerson, 1986) have distinguished between eating-disordered individuals (i.e., those who meet *DSM* criteria for an eating disorder such as anorexia or bulimia) and the two types of non-eating-disordered individuals just discussed: (a) those who have no symptoms of eating disorders (i.e., asymptomatic) and (b) those who do not meet *DSM* criteria but who nevertheless display symptoms of eating disorders, such as binging once per month or taking large quantities of laxatives on an occasional basis (i.e., symptomatic). Nevertheless, just who constitutes the symptomatic group has varied between studies; many researchers have not systematically excluded women with an EDNOS diagnosis from this group, even though they have a *DSM*-defined eating disorder. In other words, many researchers have inappropriately combined, in the symptomatic group, individuals with *DSM–IV*-defined subthreshold or atypical eating disorders (i.e., EDNOS, such as subthreshold bulimia) and individuals with non-*DSM*-classifiable clusters of symptoms (i.e., no classifiable disorder but some symptoms, such as binging twice per year). Perhaps most important, the distinctions among eating disordered, symptomatic, and asymptomatic are not routinely made and have never been validated with clinical interviews.

Despite their potential ability to make differentiations among eating-disordered, symptomatic, and asymptomatic individuals, problems with operationalized *DSM* questionnaires abound. Specifically, generalization across studies is a problem for three reasons. First, with the exception of questionnaires designed by Mintz and Betz (1988); Pyle, Halvorson, Neuman, and Mitchell (1986); and Yager, Landsverk, and Edelstein (1987), rarely has the same questionnaire been used across multiple studies. Second, many publications simply state that operationalized criteria were used without specifying the items or decision rules used to operationalize criteria (e.g., see Leon, Fulkerson, Perry, & Dube, 1994; Nevo, 1985). Third, in examinations of studies in which operationalized criteria are specified, much variability between the decision rules used to arrive at diagnoses becomes apparent. An illustration has been provided by Fairburn et al. (1990), who noted that whereas Katzman and Wolchik (1984) defined "episodic eating patterns" as eight binges per month, Dykens and Gerrard (1986) considered a binge frequency of once a month to be indicative of bulimia. Another problem with the majority of these operationalized *DSM* questionnaires is that they rarely fully operationalize *DSM* criteria. Starting with the *DSM–III–R*, an attitudinal component has been included as a criterion for bulimia, yet very few questionnaires assess these attitudes (Fairburn et al., 1990). Likewise, although a range of inappropriate compensatory behaviors constitute the *DSM* criteria for bulimia (e.g., fasting, laxative abuse, enema abuse, or excessive exercise), rarely is anything but self-induced vomiting assessed (Fairburn et al., 1990).

Perhaps the most critical problem that plagues the use of self-report data (i.e., both preexisting inventories and op-

erationalized *DSM* questionnaires) in eating disorder research is a methodological one. All but a handful of studies have selected participants for eating disorder research on the basis of responses to self-report data and have not used clinical interviews to confirm whether these individuals actually had eating disorders (Fairburn et al., 1990). Thus, some eating disorder research participants may not actually have had eating disorders, and conclusions drawn could be inaccurate. As a result of this recognition, there now exist several two-stage studies in which self-report questionnaires were followed by interviews to confirm or disconfirm the diagnosis (Button & Whitehouse, 1981; Clarke & Palmer, 1983; Meadows et al., 1986; Nagelberg, Hale, & Ware, 1984; Nevo, 1985; Shefer, 1987; Whitehouse, Phil, & Button, 1988). However, as pointed out by Fairburn et al. (1990), a problem is that these studies simply include a statement noting that individuals misdiagnosed by self-report were eliminated from the study; there are only about nine studies that have included a report of the actual percentage of individuals misdiagnosed by the self-report inventory (e.g., false-positive rates) or the level of agreement (e.g., kappa) between self-report and clinical interview diagnoses (de Zwaan et al., 1993; Freeman & Henderson, 1988; Johnsone-Sabine et al., 1988; King, 1986; Pyle et al., 1986; Schotte & Stunkard, 1987; Spitzer et al., 1993; Szmukler, 1982; Williams, Schaefer, Shisslak, Gronwaldt, & Comerci, 1986). Of these nine studies, only four have been concerned with the evaluation of questionnaires that operationalize *DSM* criteria (de Zwaan et al., 1993; Pyle et al., 1986; Schotte & Stunkard, 1987; Spitzer et al., 1993). In two of these studies, the questionnaires were based on outdated *DSM* criteria, and the evaluation of the questionnaire was not the focus of the research (Pyle et al., 1986; Schotte & Stunkard, 1987). The Spitzer et al. (1993) and de Zwaan et al. (1993) studies both dealt with the evaluation of a questionnaire that operationalizes *DSM–IV* Appendix B criteria for binge-eating disorder (bulimia is also operationalized because it must be ruled out for a diagnosis of binge-eating disorder).

It is surprising that only two studies have evaluated the validity of diagnoses obtained by operationalizing *DSM–IV* criteria into a questionnaire format. Both studies provided preliminary evidence that this widely used eating disorder research methodology (i.e., operationalizing criteria into questionnaire format and using decision rules to make diagnoses) can result in valid diagnoses. Specifically, both studies found a modest level of diagnostic agreement between a questionnaire-based and interview-based diagnosis of binge-eating disorder (de Zwaan et al., 1993, reported a kappa value of .57, and Spitzer et al., 1993, reported a kappa value of .60). These results are interesting in light of a study by Fairburn and Beglin (1994) reporting that, in a comparison of interview and self-report methods for assessing the individual features of eating disorders, the two performed similarly with respect to the assessment of unambiguous behavioral features (e.g., self-induced vomiting), whereas discrepancies emerged when more complex features (e.g., binge eating and concerns about body shape) were assessed. More data are clearly needed to determine the reliability and

validity of diagnoses yielded by operationalizing the full range of *DSM* eating disorder criteria into questionnaire format, and this was the focus of our research.

Operationalized *DSM–IV* Questionnaire Development

Choice of Base Questionnaire

Because the methodology of operationalizing *DSM* criteria into self-report questionnaire format is common in the eating disorder field, several such operationalized questionnaires exist. Hence, rather than invent such a questionnaire, we chose to modify a preexisting one. The questionnaire we chose to revise was the Weight Management Questionnaire (WMQ; Mintz & Betz, 1988, *DSM–III–R* revision of Ousley, 1986, *DSM–III* questionnaire). Although no version of the WMQ has been validated by clinical interviews or clinician judgments, several studies have used various forms of the WMQ to study *DSM–III*- and *DSM–III–R*-defined groups such as bulimics, bingers, and purgers (e.g., Kaminski & McNamara, 1996; Mintz, 1989; Mintz & Betz, 1988; O'Halloran, 1989; Scarano, 1991; Smithies, 1989).

The WMQ was used as a base for several reasons. First, it is one of the few operationalized questionnaires to be used by multiple researchers. In addition, it is one of the few existing operationalized *DSM* questionnaires that defines several eating behavior groups simultaneously (e.g., bulimics, bingers, purgers, and normals). Indeed, the WMQ was recommended in a recent review by Scarano and Kalodner-Martin (1994) as the ideal method with which to operationalize a range of *DSM–III–R*-defined eating disorders. Scarano and Kalodner-Martin (1994) noted that several researchers have used the WMQ, and they advocated that more do so because it would facilitate comparison across studies.

Revision of the WMQ

Items. Earlier versions of the WMQ were developed via the format of earlier eating disorder researchers (e.g., Fairburn & Cooper, 1983; Katzman & Wolchik, 1984) in terms of inclusion of operationalized versions of *DSM–III* and *DSM–III–R* criteria (see detailed description by Mintz & Betz, 1988). For the earlier versions, content validity was established by expert judgments. The development of this *DSM–IV* revision followed a similar methodology. Items from previous versions of the WMQ were used when *DSM–IV* criteria were identical to *DSM–III–R* criteria (e.g., for bulimia, the criteria concerning frequency of binge-eating episodes remained the same from the *DSM–III–R* to the *DSM–IV*). When *DSM–III–R* and *DSM–IV* criteria differed, new items were added or old items were revised (e.g., for bulimia, a frequency requirement for inappropriate compensatory behaviors was added to the *DSM–IV*, and hence questions assessing frequency were added to the WMQ to reflect this new criterion). Seven eating disorder experts

then examined items for content validity. Items were modified in accordance with the suggestions of the experts, who then examined the items a second time and agreed that they accurately reflected *DSM–IV* criteria. Ten graduate and undergraduate students then completed a version containing these items and were asked to comment on instruction understandability and item clarity. Minor grammatical changes were made in response to comments, and an additional five graduate and undergraduate students completed the questionnaire. All judged it to be understandable and clear.

Decision rules. Previous WMQ decision rules (previously termed *inclusion criteria;* see Mintz, 1989; Mintz & Betz, 1988) were modified to reflect changes in the *DSM–IV.* Eating disorder experts examined decision rules and provided an assessment indicating that the rules adequately reflected *DSM–IV* criteria and diagnoses.

Description of Revised Instrument

The operationalized *DSM–IV* questionnaire that resulted from the revisions just described was renamed the Questionnaire for Eating Disorder Diagnoses (Q-EDD). The Q-EDD is a self-report questionnaire that contains 50 questions and requires approximately 5 to 10 min to complete. The Q-EDD yields both frequency data for individual behaviors (e.g., self-induced vomiting) and categorical labels (e.g., eating disordered and non–eating disordered). Categorical labels are generated by a scoring manual that consists of flowchart decision rules, in which items or combinations of items are dichotomously scored (yes or no) in terms of meeting or not meeting individual *DSM–IV* criteria; individual criteria are then combined with additional decision rules into an assessment of meeting or not meeting all criteria for a specific diagnosis or category. Table 1 contains sample Q-EDD items and corresponding decision rules.

On the basis of decision rules, respondents are placed into diagnostic categories. At the most general level are the diagnostic categories of non–eating disordered and eating disordered, each of which is composed of more specific categories. The non-eating-disordered category is composed of two other categories: asymptomatic (i.e., no eating disorder symptoms) and symptomatic (i.e., some eating disorder symptoms but no *DSM–IV* diagnosis). The eating disorder category is composed of six specific diagnostic categories: two reflecting the *DSM–IV* diagnoses of bulimia and anorexia (which can be further broken down to reflect the *DSM–IV* subtypes listed in the note to Table 1) and four reflecting the *DSM–IV* EDNOS descriptions of subthreshold bulimia, menstruating anorexia, nonbinging bulimia, and binge-eating disorder. Although (as specified in the note to Table 2) the *DSM–IV* contains six EDNOS examples, only four were operationalized by the Q-EDD (normal-weight anorexia was not studied because of difficulties in operationalizing significant weight loss into a pencil-and-paper measure, and "chew-spitting" was not operationalized because preliminary interviews indicated that the few respon-

dents who endorsed only such an item indicated that they spit out food that did not taste good).

In terms of specific decision rules, respondents need to meet the full diagnostic criteria for any one of the six eating disorders to be placed in the eating-disordered category (i.e., they need to receive a Q-EDD diagnosis of anorexia or bulimia, or meet the operationalized criteria for one of the four EDNOS categories). Respondents need to meet all of the *DSM–IV* criteria for anorexia and bulimia to receive these same diagnoses on the Q-EDD. On the other hand, even though *DSM–IV* descriptions guided decision rules, some degree of subjectivity was entailed in defining EDNOS groups as a result of the term "regular use" in the *DSM–IV* EDNOS descriptions, with no clear operationalization of regular use being provided (see Spitzer et al., 1992, 1993, and *DSM–IV* Appendix B for more detailed discussion of this issue). For the purposes of the Q-EDD, decision rules for classification into one of the four EDNOS groups were based on rather strict operational definitions, as described in Table 2. For classification into the asymptomatic category, decision rules specify that individuals need to respond negatively to all behaviors constituting *DSM–IV* eating disorder criteria and also need to respond negatively to the use of strict dieting and appetite control pills as a means to control weight. Although these latter two items are not part of the *DSM–IV* criteria, they are behaviors often engaged in by eating-disordered individuals or individuals at risk for eating disorders (Lachenmeyer & Muni-Brander, 1988; Moreno & Thelen, 1993). Asymptomatics who are either severely underweight (i.e., anorexic weight) or grossly obese are considered "red flag" asymptomatics to signify that their weight category raises suspicions that perhaps they are not truly asymptomatic (i.e., perhaps they are anorexics or binge eaters who are not responding honestly). Finally, decision rules specify that the classification of symptomatic is given when an individual does not meet the criteria for any of the *DSM–IV* eating disorders (i.e., anorexia, bulimia, or any of the four EDNOS descriptions) but is not asymptomatic. In other words, these would be individuals who report engaging in eating-disordered behaviors that do not meet criteria for any *DSM–IV* eating disorder diagnosis. These individuals might theoretically be considered "at risk" for eating disorders, yet they are not currently identified by any preexisting eating disorder instrument.

Study 1

This was the first in a series of three studies aimed at evaluating the reliability and validity of the Q-EDD. This study represents the first attempt in the field of eating disorders to provide a comprehensive assessment of the widely used methodology of operationalizing the *DSM* into a questionnaire format. Given the gaps in eating disorder assessment discussed previously, we were interested in examining the ability of the Q-EDD to make diagnostic differentiations that no current inventory is capable of providing. Specifically, we were interested in the ability of the

Q-EDD to differentiate (a) between *DSM–IV* eating-disordered (bulimic, anorexic, and EDNOS combined) and non-eating-disordered (i.e., no *DSM–IV* eating disorder diagnoses) individuals (b) among eating-disordered, symptomatic (i.e., some eating disorder symptoms but no actual *DSM–IV* diagnosis), and asymptomatic (i.e., no eating disorder symptoms) individuals, and (c) between individuals with anorexia and individuals with bulimia.

Two exploratory issues were also addressed. The first was the Q-EDD's ability to make reliable and valid differentiations among the six eating disorder categories (i.e., its ability to differentiate among bulimia, anorexia, menstruating anorexia, subthreshold bulimia, nonbinging bulimia, and binge-eating disorder). The second involved further describing and studying symptomatic individuals (i.e., the generation of labels with which to describe these individuals and the exploration of whether these labels could be used to make reliable and valid distinctions). The former purpose was considered exploratory because no other research has examined the ability of a questionnaire to make such fine diagnostic distinctions, despite the clear need for a questionnaire able to do so (Williamson et al., 1995). The latter purpose was considered exploratory because this is the first study to specifically target symptomatic individuals in general, let alone to describe subtypes of such individuals. The aim of both exploratory investigations was to lay the foundation for future research.

This study was conducted with a nonclinical sample of college women, because eating disorders and eating disorder symptoms have been found to be quite prevalent in such samples (e.g., Mintz & Betz, 1988). Criterion validity, the extent to which a measure corresponds to an accurate, independent, or external indicator of the same attribute (Kaplan & Saccuzzo, 1993; Walsh & Betz, 1985), was assessed by an examination of the correspondence between diagnoses yielded by the Q-EDD and those yielded by clinical interviews. Convergent validity was assessed by an examination of the correspondence between Q-EDD diagnoses and scores on the BULIT-R and the EAT (e.g., convergent validity would be demonstrated if individuals diagnosed as bulimic by the Q-EDD scored in the bulimic range on the BULIT-R). Incremental validity, the extent to which a new assessment tool improves the accuracy of diagnoses above that of existing tests (Walsh & Betz, 1985), was examined by comparing the level of agreement between Q-EDD diagnoses and clinical interview diagnoses with the level of agreement between preexisting inventory diagnoses (e.g., BULIT-R) and clinical interview diagnoses (i.e., Did the Q-EDD diagnose more or less accurately than existing tests?). Test–retest reliability over 1 to 3 months was also assessed (i.e., To what degree did Q-EDD diagnoses change over time?). Finally, because the Q-EDD is scored by decision rules, it was important to assess whether independent individuals scoring Q-EDDs would arrive at the same diagnostic categories; thus, interscorer agreement was also examined. All of these forms of reliability and validity were examined for our three primary diagnostic differentiations, whereas only select types were examined for our two exploratory questions.

Method

Participants. Participants included 136 women from a large midwestern public university, 11 female counseling psychology graduate students who conducted clinical interviews, and 3 female psychology undergraduate students who rated tapes of these interviews. The 136 women interviewed ranged in age from 18 to 41 years ($M = 19.04$, $SD = 2.69$). The majority (88%) were Caucasian; 5% were African American, 3% were Hispanic–Latino–Mexican American, 2% were Asian American, 1% were Native American, and 1% either stated that they were of another ethnicity or did not report their ethnicity. Also, most of these participants were freshmen (79%); 14%, 5%, and 2% were sophomores, juniors, and seniors, respectively. On the basis of the results of structured clinical interviews, the 136 participants included 33 eating-disordered women (1 anorexic, 9 bulimics, 4 menstruating anorexics, 12 subthreshold bulimics, 4 nonbinging bulimics, and 3 binge eaters) and 103 non-eating-disordered women (20 symptomatic and 83 asymptomatic). Of the 11 counseling psychology graduate students who conducted the interviews, 8 were Caucasian, 1 was Hispanic, and 1 was Asian American; their mean age was 29.81 years. The 3 undergraduates who rated tapes were all Caucasian; their mean age was 20.33 years.

Instruments. The BULIT-R (Thelen et al., 1991) was used in this study to provide an indication of the convergent and incremental validity of the Q-EDD. The BULIT-R is a widely used self-report measure of bulimia, the purpose of which is "to identify women who are most likely to be diagnosed as bulimic based on an interview" (Thelen et al., 1991, p. 123). According to a recent study (Thelen et al., 1996), the BULIT-R differentiates *DSM–IV*-diagnosed bulimic women from all other women, including asymptomatic women and women with other eating disorders and eating-disordered behaviors (in Q-EDD terms, it differentiates bulimics from a combined group consisting of all other eating disorder categories, the symptomatic category, and the asymptomatic category).

Of the 36 items included in the BULIT-R, only 28 are used to determine the final score. All 36 items are presented in a 5-point Likert scale format (1 point is given for extreme "normal" responses, and 5 points are given for extreme "bulimic" responses). Total scores, obtained by summing across the 28 items, can range from 28 to 140; 104 is the cutoff for classification as bulimic.

Thelen et al. (1991) found that the scale's 2-month test–retest reliability was .95 and that its internal consistency reliability was .97. In addition, when correlating BULIT-R scores and group membership (bulimic vs. nonbulimic) based on rater judgments, Thelen et al. (1991) reported a correlation coefficient of .62 ($p < .001$). A recent study found BULIT-R scores to be highly predictive of a *DSM–IV* diagnosis of bulimia, as indicated by clinical interviews and clinician judgments; in correlation analyses of BULIT-R scores and group membership (bulimic vs. nonbulimic), a correlation coefficient of .73 ($p < .0001$) was found (Thelen et al., 1996). Thelen et al. (1991) also reported a correlation of .85 between the BULIT-R and Hawkins and Clement's (1980) Binge Scale.

The EAT (Garner & Garfinkel, 1979) was also used in this study to provide an indication of the convergent and incremental validity of the Q-EDD. The EAT is a widely used self-report measure, the purpose of which is to provide "an objective and valid index of symptoms frequently observed in anorexia nervosa" (Garner & Garfinkel, 1979, p. 276). Because the EAT was developed with two samples of clinical anorexics and normal controls, it was originally designed to differentiate anorexics and non-eating-disordered individuals. As noted earlier, even though the authors

of the EAT (Garner et al., 1982) consider it to be a measure of general disordered eating in nonclinical samples, it is still widely used as a measure of anorexia in research conducted with such samples.

The EAT's 40 items are presented in a 6-point forced-choice Likert scale format. Total scores are obtained by giving the most "symptomatic" response a score of 3, the next most extreme response a score of 2, and the adjacent less extreme response a score of 1; no score is given for "nonanorexic" answers (Garner & Garfinkel, 1979, pp. 274, 278). A score of 30 is the cutoff for classification as anorexic. Internal consistency reliabilities were found by Garner and Garfinkel (1979) to range from .79 to .94.

Clinical interview. The structured clinical interview followed the format of the Structured Clinical Interview for Axis I DSM–IV Disorders (SCID) for Module H (Eating Disorders; First, Spitzer, Gibbon, Gibbon, & Williams, 1994). The SCID is the only interview measure "to correspond exactly to the criteria established by the DSM" (Pike et al., 1995, p. 321). The SCID thus "provides the essential information for discriminating among the eating disorders" (Pike et al., 1995, p. 323). Test–retest reliability of the bulimia nervosa section of the *DSM–III–R*-based SCID has been found to range from .82 to .90 (Pike et al., 1995). No study has specifically examined the interrater reliability of SCID eating disorder diagnoses (Pike et al., 1995); however, interrater reliabilities of SCID diagnoses are generally found to be high (Skre, Onstad, Torgersen, & Kringlen, 1991). Another advantage of the SCID is that the interviewer may add or restructure items when participants' responses are equivocal (Pike et al., 1995); for this study, interviewers were trained to ask follow-up questions to

ensure sufficient information with which to make diagnostic judgments (e.g., questions concerning food eaten during a typical binge, to assess whether the amount eaten was truly more than the average person would eat during a similar time period, and questions specifically confirming frequency of binges and compensatory behaviors).

Procedures. The Q-EDD was administered to approximately 1,400 college women across two semesters as part of a mass testing of introductory psychology students. Participants were placed in eating disorder categories on the basis of their responses on the Q-EDD. Q-EDDs were scored by undergraduate research assistants, who used the scoring manual described previously to categorize respondents into the main diagnostic categories (eating disordered and non–eating disordered), eating disorder subcategories (anorexia, bulimia, subthreshold bulimia, menstruating anorexia, binge-eating disorder, and nonbinging bulimia), and non-eating-disordered subcategories (asymptomatic and symptomatic).

For one of our exploratory purposes (i.e., the study of symptomatic women), the Q-EDD of every participant classified into the symptomatic category was visually inspected by Laurie B. Mintz, Amy M. Mulholland, and Paxton A. Schneider. On the basis of this visual inspection, descriptive labels were generated. These descriptive labels were based on eating disorder nomenclature. For example, a person who reported vomiting once a month but not binging was described as a subthreshold nonbinging bulimic, whereas a person who met all of the behavioral criteria for bulimia but did not meet the attitudinal criteria was described as a behavioral bulimic. Table 3 contains a complete listing of these descriptive labels. After this list had been generated, it was given

Table 3
Descriptive Labels (Subtypes) of Symptomatic Women

Label	Description
Low-weight anorexia	Body mass index of 17.6–19.0 and meets all other criteria for anorexia
Nonnormal-weight nonbinging bulimia	Meets all criteria for nonbinging bulimia but is in a weight category other than normal[a]
Subthreshold nonbinging bulimia	Any weight category, no binges, compensates (i.e., fasting, vomiting) but not at a high enough frequency to be classified as having nonbinging bulimia
Subthreshold binge-eating disorder	All criteria for binge-eating disorder but not at a high enough frequency
Binge dieter	Binges and compensates by strict dieting (no other compensatory behaviors such as fasting or vomiting)
Behavioral bulimia	Meets all criteria for bulimia, including frequency, but reports feeling in control during binges or that self-esteem is not unduly influenced by weight or body shape
Subthreshold behavioral bulimia	Meets all criteria for bulimia except frequency and reports feeling in control during binges or that self-esteem is not unduly influenced by weight or body shape
Chronic dieter	Does not binge, uses strict dieting or appetite control pills but no inappropriate compensatory behavior (i.e., fasting, vomiting, excessive exercise, laxatives)

Note. In the scoring manual, a category of "other" is also included for individuals who do not fall into any of the listed categories, and the scorer is asked to give descriptive labels and to describe behaviors.

[a] Possible weight categories based on body mass index are low weight, overweight, moderately obese, and grossly obese. Severely underweight individuals with the behavior would be captured by anorexia criteria.

to undergraduate research assistants, who were asked to examine all Q-EDDs in the symptomatic category and to choose the descriptive label that best fit each one.

As a means of determining interscorer agreement in scoring Q-EDDs, the Q-EDDs of 50 of the 136 participants were scored by two separate undergraduate research assistants. Both scorers were asked to use the scoring manual for the main diagnoses, as well as to select a descriptive label for each Q-EDD falling in the symptomatic category.

All respondents whose mass-testing responses placed them in the eating disorder category and a random sample of non-eating-disordered women were called and asked to participate in the study for course credit; approximately 76% of the participants contacted agreed to take part (the majority of those who declined did so because they did not need the course credit). The time period between mass testing and participation ranged from 1 to 3 months. Participation consisted of completing the Q-EDD, the BULIT-R, and the EAT (the order of administration was counterbalanced), as well as the structured diagnostic interview. This interview was audiotaped and subsequently listened to and rated by one of the three undergraduate tape raters. Although these tape raters also scored Q-EDDs, the study was designed so that each participant's Q-EDD was scored by an individual other than the one who interviewed her or rated her tape (i.e., no overlap among Q-EDD scorer, interviewer, and tape rater).

Before the study, the interviewers and tape raters participated in a training session, conducted by Laurie B. Mintz, consisting of didactic information on the DSM–IV, training on the interview protocol, and role-plays of interviews and ratings. A second training session was provided at the midpoint of the study, in which interviewers and tape raters had the opportunity to listen to each other's tapes and to provide feedback on interviews and ratings. Laurie B. Mintz also reviewed the first several interviews and diagnostic judgments with raters and interviewers and provided feedback. Both interviewers and raters were unaware of the participants' diagnostic category as assessed by the Q-EDD at prescreening. The average interview was approximately 30 min long. After completing the interview, all participants were debriefed and provided with counseling resources for eating disorders.

On the basis of the interview, the interviewer judged whether or not each participant met each of the DSM–IV criteria for eating disorders and placed each participant into a diagnostic category. The diagnostic categories were the same as those yielded by the Q-EDD. In addition, for our exploratory purpose, interviewers were asked to choose the descriptive label that best fit each participant they diagnosed as symptomatic. For each interview, a second diagnosis was provided by one of the three undergraduate tape raters who listened to audiotapes of the interviews. A third independent rater was used in cases of disagreement. For the 136 interviews, a third rater was needed eight times, and two of these occasions concerned only disagreement over symptomatic descriptive labels. Final diagnoses were determined by the judgments of the interviewer and rater or, in the cases of disagreement, the diagnoses agreed on by two of the three. Interrater agreement was very high for two of our three primary diagnostic differentiations: (a) 99% (κ = .98) for the differentiation between eating disordered and non–eating disordered and (b) 97% (κ = .95) for the differentiation of eating disordered, symptomatic, and asymptomatic. Because we had only 1 interview-defined anorexic, interrater reliability for the diagnostic differentiation of anorexia and bulimia was meaningless (although it is interesting to note that there was 100% agreement for this differentiation). Interrater agreement was also high for our exploratory differentiation of (a) the six eating disorders from one another (94%, κ = .92) and (b) symptomatic

subtypes (89%; the kappa value could not be calculated because rows and columns were not equal).

Results

Because we had only 1 interview-defined anorexic, analyses concerning the diagnostic differentiation of anorexia and bulimia could not be conducted, except as noted subsequently. Instead, descriptive information on the 1 anorexic is provided as appropriate.

Convergent validity. Convergent validity was examined via t tests and analyses of variance (ANOVAs); BULIT-R scores, EAT scores, and Q-EDD categories were used. For these analyses, the category yielded by the second Q-EDD administration was used because this administration corresponded to the administration of the BULIT-R and the EAT.

The BULIT-R is valid only for the differentiation between DSM–IV bulimics and nonbulimics (i.e., all other eating disorders and non–eating disordered combined), and hence convergent validity was examined through the use of these categories. The BULIT-R scores of the Q-EDD-defined bulimics (M = 108.67) and nonbulimics (M = 53.89) were significantly different, $t(133)$ = 6.67, p < .0001, with the mean score of the Q-EDD-defined bulimics falling in the bulimic range (i.e., more than 104) and the mean score of the Q-EDD-defined nonbulimics falling in the nonbulimic range.

The EAT is designed to differentiate anorexics and non-eating-disordered individuals (i.e., normal controls). The anorexic, menstruating anorexic, and non-eating-disordered categories were used in examining convergent validity. Although the EAT is designed for anorexia, it was thought that combining anorexia and menstruating anorexia was legitimate because (a) menstruating anorexia is an atypical form of anorexia, with the only difference between them being the presence or absence of menstruation, and (b) the EAT contains only one question on menstruation. The EAT scores of the Q-EDD-defined anorexics–menstruating anorexics (M = 30.50) and non-eating-disordered participants (M = 13.20) were significantly different, $t(104)$ = 3.65, p < .001, the mean score of Q-EDD-defined anorexics–menstruating anorexics falling in the anorexic range (i.e., more than 30) and the mean score of Q-EDD-defined non-eating-disordered participants falling in the nonanorexic range.

Because its authors (Garner et al., 1982) stated that, in nonclinical samples, the EAT is a measure of disturbed eating attitudes and behaviors (i.e., rather than just anorexia), convergent validity of the Q-EDD was also examined in an ANOVA involving Q-EDD category (eating disordered, symptomatic, or asymptomatic) as the independent variable and EAT score as the dependent variable. This ANOVA was significant, $F(2, 134)$ = 99.65, p < .0001. Post hoc contrast analyses indicated that the mean EAT scores of the three Q-EDD-defined groups differed significantly from one another (p < .001). Mean EAT scores were as follows: eating-disordered group, 42.96; symptomatic group, 21.76; and asymptomatic group, 10.42.

Criterion validity. Criterion validity was examined through (a) the level of agreement between Q-EDD categories and categories determined by the structured clinical interview (i.e., the final diagnosis based on interviewer and rater judgments) and (b) an evaluation of how well the Q-EDD categorized respondents into correct diagnostic categories. For level of agreement, the kappa statistic was used. Although often thought of as a measure of interrater agreement, kappa is also a general measure of agreement that can be used with nominal data to examine validity and test–retest reliability (see Crews & Sher, 1992; Liebetrau, 1983). Accuracy rates were used in evaluating how well the Q-EDD categorized respondents into accurate diagnostic categories (i.e., those determined by the interview); an accuracy rate is the percentage of cases in which a test (in this case, the Q-EDD) produces an accurate diagnosis (Kaplan & Saccuzzo, 1993). Along with accuracy rates, we also calculated, for dichotomous diagnostic differentiations (e.g., eating disordered or non–eating disordered), false-negative rates, false-positive rates, sensitivity, specificity, positive predictive power, and negative predictive power. These are widely used statistics for evaluating the predictive ability of inventories that yield positive–negative (e.g., diagnosis–no diagnosis) dichotomous data. Specifically, sensitivity, the conditional probability of the presence of a risk factor (in this case, Q-EDD category) given the presence of the disorder (as determined by interviewer), is the true-positive rate. Specificity, the conditional probability of not having the risk factor if one does not have the disorder, is the true-negative rate. Positive predictive power, the conditional probability of having the disorder given the presence of the risk factor, is the percentage of time one would be correct in predicting the presence of the disorder when using the inventory. Negative predictive power is the conditional probability of not having the disorder given the lack of the risk factor (for more information and for computational formulas, see Kaplan & Saccuzzo, 1993; Meehl & Rosen, 1955; Widiger, Hurt, Frances, Clarkin, & Gilmore, 1984).

With respect to the eating-disordered versus non-eating-disordered diagnostic differentiation, 1 eating-disordered participant was misclassified by the Q-EDD as non–eating disordered, and 2 non-eating-disordered participants were misclassified as eating disordered; all other participants were correctly classified. Hence, the accuracy rate was 98%. Predictive ability statistics were as follows: false-negative rate, .03; false-positive rate, .02; sensitivity, .97; specificity, .98; positive predictive power, .94; and negative predictive power, .99. The kappa value was .94.

Results for the eating-disordered, symptomatic, and asymptomatic diagnostic differentiation overlapped somewhat with those just presented because asymptomatic and symptomatic are non-eating-disordered subcategories. Specifically, the 1 eating-disordered participant misdiagnosed by the Q-EDD as non–eating disordered was misdiagnosed as symptomatic. Of the 2 non-eating-disordered participants who were misdiagnosed by the Q-EDD as eating disordered, 1 was symptomatic and 1 was asymptomatic. In addition, 3 interview-diagnosed symptomatics were misdiagnosed as

asymptomatic by the Q-EDD, and 8 interview-diagnosed asymptomatics were misdiagnosed as symptomatic by the Q-EDD. Hence, the accuracy rate for this differentiation was .90 ($\kappa = .82$).

Our third primary diagnostic differentiation (anorexia vs. bulimia) could not be examined. Nevertheless, it should be noted that our 1 interview-defined anorexic was misdiagnosed by the Q-EDD as bulimic because her body mass index (BMI) of 17.7 exceeded the scoring manual BMI cutoff of 17.5. In examinations of anorexia and bulimia, it is also interesting that subtype was never missed.

Incremental validity. The incremental validity of the Q-EDD would be defined as the extent to which it improves diagnostic accuracy above that of existing tests such as the BULIT-R and the EAT. Incremental validity thus involved a comparison of existing tests (e.g., BULIT-R) and the Q-EDD with respect to how well they categorized respondents into correct diagnostic categories (i.e., a comparison of their criterion validity). In this study, both the Q-EDD and the BULIT-R correctly diagnosed 7 bulimics; the Q-EDD correctly diagnosed 125 nonbulimics, and the BULIT-R correctly diagnosed 120 nonbulimics. Both the Q-EDD and the BULIT-R misdiagnosed 2 bulimics as nonbulimic. The Q-EDD misdiagnosed 2 nonbulimics as bulimic, whereas the BULIT-R made this misdiagnosis six times. Thus, the performance of the BULIT-R and Q-EDD in differentiating bulimics and nonbulimics was as follows (statistics associated with the Q-EDD are presented first, followed by those for the BULIT-R): accuracy rate, .97 and .94; false-negative rate, .22 and .22; false-positive rate, .02 and .05; sensitivity, .78 and .78; specificity, .98 and .95; positive predictive power, .78 and .54; negative predictive power, .98 and .98; and kappa value, .76 and .61. Hence, the Q-EDD and the BULIT-R were roughly equivalent in all respects except for positive predictive power (considered most important by Widiger et al., 1984); the Q-EDD was correct at predicting bulimia 78% of the time, whereas the BULIT-R was correct 54% of the time. The incremental validity of the Q-EDD in comparison with that of the EAT was not examined, again because there was only one interview-defined anorexic.

Reliability. Test–retest reliability was calculated, via contingency tables and kappa values, in terms of change in Q-EDD categories (i.e., diagnoses) from the first administration to the second administration of the Q-EDD (as noted in the *Procedure* section, this period ranged from 1 to 3 months). Kappa values and specifics regarding changes (total changes in both directions) were as follows: (a) eating-disordered and non-eating-disordered groups $\kappa = .64$ (19 changes between the eating-disordered and non-eating-disordered categories); and (b) eating-disordered, symptomatic, and asymptomatic groups $\kappa = .54$ (14 changes between the asymptomatic and symptomatic categories, 13 changes between the symptomatic and eating-disordered categories, and 6 changes between the asymptomatic and eating-disordered categories).

Interscorer agreement. Fifty randomly selected Q-EDDs were scored by two separate scorers. There was 100% agreement ($\kappa = 1.00$) between the scorers for the diagnostic differ-

entiations of (a) eating disordered and non–eating disordered and (b) eating disordered, symptomatic, and asymptomatic. No randomly selected Q-EDDs were scored as anorexic, so no data were available on interscorer agreement for the differentiation of anorexia and bulimia.

Exploratory Analyses

Differentiations among eating disorder groups. Our first exploratory question concerned the ability of the Q-EDD to make reliable and valid differentiations among the six eating disorder groups. In terms of how well the Q-EDD categorized respondents into correct diagnostic categories (i.e., criterion validity), 2 of the 12 subthreshold bulimics (as determined by the interview) were misclassified by the Q-EDD as bulimic. One interview-defined bulimic was misclassified as a nonbinging bulimic. As already noted, 1 interview-defined anorexic was misclassified as bulimic by the Q-EDD. None of the menstruating anorexics, binge eaters, or nonbinging bulimics were misclassified. The accuracy rate for the differentiation of bulimics, subthreshold bulimics, anorexics, menstruating anorexics, binge eaters, and nonbinging bulimics was 88% (κ = .84). When bulimia and subthreshold bulimia were combined, the accuracy rate increased to 94% (κ = .89). There was 100% agreement between the two scorers of the Q-EDD for the differentiation of all eating disorder Q-EDDs randomly selected. The test–retest kappa value was .46 (seven changes between subthreshold bulimia and bulimia, one between subthreshold bulimia and nonbinging bulimia, one between binge-eating disorder and bulimia, one between binge-eating disorder and subthreshold bulimia, and one between menstruating anorexia and bulimia). When bulimia and subthreshold bulimia were combined, the test–retest kappa value increased to .75.

Validity of symptomatic descriptive labels. Our second exploratory question concerned the use of symptomatic descriptive labels to make reliable and valid distinctions. In terms of criterion validity, in the 16 cases in which a descriptive label was generated for both the Q-EDD and the clinical interview (i.e., diagnosed as symptomatic by both), there was a 69% agreement between them. In terms of interscorer agreement, there was 100% agreement between the two scorers of the Q-EDD in terms of symptomatic descriptive labels. As noted earlier (see *Procedure* section), there was also an 89% agreement between interviewers and tape raters concerning symptomatic labels (i.e., interrater reliability).

Additional information on symptomatic subtypes can be gleaned by examining the symptomatic subtypes involved in misses between interview-based and Q-EDD-based diagnoses for our primary diagnostic distinctions. For example, an examination of the misses between symptomatic and asymptomatic (in both directions) reveals that they all concerned three Q-EDD symptomatic subtypes and strikingly similar situations: 4 of the 11 involved "subthreshold nonbinging bulimics, restricting type" whose Q-EDD responses indicated once a month fasting; 3 involved "chronic diet-

ers"; and 4 involved "subthreshold binge eaters" for whom the discrepancy concerned whether or not what the individual ate constituted a binge (e.g., the individual indicated binging but the description of what was eaten during a binge was not judged by the interviewer to be a binge, or vice versa). The misses between eating disordered and symptomatic included 1 individual diagnosed as bulimic by the interview and as a symptomatic "binge dieter" by the Q-EDD. In addition, 1 individual was diagnosed through the interview as a symptomatic "subthreshold behavioral bulimic" and as a subthreshold bulimic by the Q-EDD; the only difference, then, is that the interviewer and rater did not judge her self-esteem to be unduly influenced by weight, whereas her Q-EDD answers indicated that she did meet this criterion. In short, when the framework of symptomatic subtypes was used, it appears that misses generally involved what might be considered related or parallel diagnostic labels.

Study 2

Study 2 was concerned with examining the convergent validity, test–retest reliability, and interscorer agreement of the Q-EDD when administered to a different sample of university women. Because the test–retest reliability assessed in Study 1 involved a rather long time period (1 to 3 months), we were interested, in Study 2, in the test–retest reliability of the Q-EDD for a shorter (i.e., 2-week) period. We were also interested in gathering more data for our exploratory foci; hence, interscorer agreement for the differentiation of eating disorder categories and symptomatic subtypes was calculated, as was test–retest reliability across the six eating disorder categories.

Method

Participants. Participants were 167 women from a medium-sized western public university. Participants' ages ranged from 18 to 51 years (M = 22.50, SD = 5.41). The majority (92%) were Caucasian; 1% were African American, 3% were Hispanic–Latino–Mexican American, 3% were Asian American, and 1% did not report ethnicity. Fifteen percent of the participants were freshmen, 23% were sophomores, 28% were juniors, 31% were seniors, and 3% were graduate or unclassified students.

Instruments. The instruments (Q-EDD, BULIT-R, and EAT) were the same as those used in Study 1.

Procedures. Faculty in departments with a high percentage of female students (i.e., anthropology, psychology, education, physical education, and nursing) were contacted and asked whether they would allow time at the end of two class sessions to collect data; faculty from all departments agreed. As an incentive, participants were entered into a raffle for a gift certificate at the university bookstore after completion of both data collection sessions. Approximately 82% of those asked to participate agreed to do so. During the first data collection, participants completed the Q-EDD, the BULIT-R, and the EAT. These instruments were presented in a counterbalanced order. Two weeks later, participants completed another Q-EDD. As in Study 1, 50 randomly selected Q-EDDs were scored by two separate scorers.

Results

On the basis of their responses to the first administration of the Q-EDD, 11 participants were categorized as eating disordered (1 anorexic, 1 menstruating anorexic, 1 bulimic, 4 subthreshold bulimics, 2 nonbinging bulimics, and 2 binge eaters), 46 were categorized as symptomatic, and 110 were categorized as asymptomatic.

Convergent validity. We had planned to examine convergent validity in Study 2 as we had in Study 1 (i.e., via *t* tests and ANOVAs). However, because there was only 1 Q-EDD-defined bulimic, convergent validity between the Q-EDD bulimic and nonbulimic categories and BULIT-R scores could not be calculated. Likewise, as a result of the small number of Q-EDD-defined anorexics, convergent validity between the Q-EDD anorexic and non-eating-disordered categories and EAT scores could not be calculated. An ANOVA with Q-EDD category (eating disordered, symptomatic, or asymptomatic) as the independent variable and EAT score as the dependent variable was significant, $F(2, 158) = 24.13, p < .0001$. Post hoc contrast analyses indicated that the mean EAT score of Q-EDD-defined asymptomatic participants (11.71) differed significantly from that of both Q-EDD-defined symptomatic (21.49) and eating-disordered (24.44) participants; there were no significant differences between symptomatic and eating-disordered mean EAT scores.

Reliability. Two-week test–retest reliability was calculated in terms of changes from the first administration to the second administration of the Q-EDD. As in Study 1, test–retest reliability was calculated via contingency tables and kappa values. Kappa values and specifics regarding changes (total changes in both directions) were as follows: (a) eating-disordered and non-eating-disordered groups $\kappa = .94$ (one change between the eating-disordered and non-eating-disordered categories), and (b) eating-disordered, symptomatic, and asymptomatic groups $\kappa = .85$ (one change between the symptomatic and eating-disordered categories and eight changes between the asymptomatic and symptomatic categories). There were no changes between anorexia and bulimia, but these results must be viewed with extreme caution given that there was only 1 participant in each category.

Interscorer agreement. Fifty randomly selected Q-EDDs were scored by two separate scorers. There was 100% agreement ($\kappa = 1.00$) between them for our diagnostic differentiations of (a) eating disordered and non–eating disordered and (b) eating disordered, symptomatic, and asymptomatic. No randomly selected Q-EDDs were scored as anorexic, so no data were available on interscorer agreement for the differentiation of anorexia and bulimia.

Exploratory Analyses

Differentiations among eating disorder groups. There was 100% agreement between the two scorers of the Q-EDD for the differentiation of all eating disorder Q-EDDs randomly selected. The test–retest kappa value was 1.00 (no changes).

Validity of symptomatic descriptive labels. There was 98% agreement between the two scorers of the Q-EDD in terms of symptomatic descriptive labels.

Study 3

Because Study 2 did not examine criterion validity and because Study 1 examined it in a nonclinical sample, a study assessing criterion validity in a clinical sample was needed. The purpose of Study 3 was to assess the criterion validity of Q-EDD diagnoses in a clinical sample of eating-disordered women by means of accuracy rates and level of agreement between Q-EDD diagnoses and clinician diagnoses. Study 1 indicated that, in a nonclinical sample, the Q-EDD demonstrated good criterion validity (i.e., high accuracy rates and level of agreement between Q-EDD and clinical interview diagnoses) for two of our three primary diagnostic differentiations (i.e., between eating disordered and nondisordered and among eating disordered, symptomatic, and asymptomatic); we were thus interested in determining whether this good criterion validity would hold for a clinical sample. In Study 1, as a result of a small sample size, we were not able to evaluate our third primary diagnostic differentiation of anorexia and bulimia; hence, we were especially interested in the differentiation of anorexia and bulimia in a clinical sample. We also continued to gather data for our exploratory foci by examining (a) the criterion validity for the differentiation of the six eating-disordered groups and (b) symptomatic subtypes involved in misses between the eating-disordered and symptomatic categories.

Method

Participants. Participants were 37 women recruited by referral from therapists in several midwestern and southern states. Participants ranged in age from 15 to 44 years ($M = 24.68$, $SD = 7.59$). The majority (97%) were Caucasian; 3% were African American. Five percent were in high school, 49% were in college, 43% were employed, and 3% were in the midst of employment–school transitions. According to clinician judgments, there were 16 anorexics, 15 bulimics, 2 subthreshold bulimics, 1 nonbinging bulimic, and 3 binge eaters.

Instruments. The Q-EDD was used.

Procedures. Q-EDDs were mailed to the directors of several midwestern and southern eating disorder clinics and the director of one large midwestern university counseling center, along with clinician rating sheets listing *DSM–IV* criteria for bulimia and anorexia, the four EDNOS descriptions of menstruating anorexia, subthreshold bulimia, nonbinging bulimia, and binge-eating disorder. Clinic directors distributed Q-EDDs and clinician rating sheets to therapists, who solicited clients' participation and filled out rating sheets indicating clients' diagnoses.

Results

Because all participants were diagnosed by clinicians as eating disordered (i.e., there were no symptomatic or asymptomatic participants), kappa values could not be calculated (i.e., row values did not equal column values in

contingency tables). Likewise, for the differentiation of eating disordered and non–eating disordered, only the calculations of false-negative rates and sensitivity were meaningful (i.e., false-positive rate, specificity, positive predictive power, and negative predictive power were misleading as a result of zeros in some of the cells on which calculations were based), with overall accuracy rate equal to sensitivity in these cases. This was also true for the differentiation of anorexia and bulimia. For the differentiation of eating disordered, symptomatic, and asymptomatic, only accuracy rate could be calculated.

Criterion validity was examined through the correspondence between Q-EDD diagnoses and diagnoses reported by clinicians. In terms of the differentiation of eating disordered and non–eating disordered, 29 individuals were diagnosed by both the Q-EDD and the clinician as eating disordered, whereas 8 individuals were diagnosed by the Q-EDD as non–eating disordered and by the clinician as eating disordered (hence, a false-negative rate of 22% and sensitivity and accuracy rates of 78%). In terms of the differentiation of eating disordered, symptomatic, and asymptomatic, 6 individuals were diagnosed as eating disordered by clinicians and as symptomatic by the Q-EDD, whereas 2 individuals were diagnosed as eating disordered by clinicians and as asymptomatic by the Q-EDD (again, an accuracy rate of 78%). Of note was the fact that both Q-EDD asymptomatics were "red-flag" asymptomatics: 1 was a severely underweight asymptomatic diagnosed as anorexic by the clinician, and 1 was a grossly obese asymptomatic diagnosed as a binge eater by the clinician.

For the differentiation of anorexia and bulimia, the kappa value was 1.00. In other words, there were no misses between the categories of anorexia and bulimia. Thus, the sensitivity and accuracy rates were 100%, and the false-negative rate was 0%.

Exploratory Analyses

Differentiations among eating disorder groups. In terms of how well the Q-EDD categorized respondents into correct diagnostic categories (i.e., criterion validity), 3 clinician-diagnosed anorexics were diagnosed as menstruating anorexics by the Q-EDD, 1 clinician-diagnosed subthreshold bulimic was diagnosed as bulimic by the Q-EDD, 4 clinician-diagnosed bulimics were diagnosed as subthreshold bulimics by the Q-EDD, and 1 clinician-diagnosed bulimic was diagnosed as a nonbinging bulimic by the Q-EDD. Hence, the accuracy rate for this differentiation was 69%. The accuracy rate increased to 86% when bulimia was combined with subthreshold bulimia and further increased to 97% when anorexia was combined with menstruating anorexia.

Validity of symptomatic descriptive labels. Information on symptomatic labels was gleaned by examining the Q-EDD symptomatic subtypes involved in misses between clinician-based and Q-EDD-based diagnostic differentiations. Specifically, the 6 misdiagnosed Q-EDD symptomatics included 3 Q-EDD "nonnormal-weight nonbinging bu-

limics" diagnosed as nonbinging bulimic, bulimic, or anorexic by clinicians; 1 Q-EDD "behavioral bulimic" diagnosed as subthreshold bulimic by the clinician; 1 Q-EDD "subthreshold binge eater" diagnosed as a binge eater by the clinician; and 1 "subthreshold exercise bulimic" diagnosed as bulimic (restricting type) by the clinician. Again, when the framework of symptomatic subtypes was used, it appears that misses generally involved what might be considered closely related diagnostic labels.

General Discussion

Strong support was obtained for the psychometric properties of the Q-EDD. For example, convergent validity was demonstrated by significant correspondence between Q-EDD diagnoses and scores on the BULIT-R and the EAT. Test–retest reliabilities found were within the expected range, given that eating disorder symptoms are not temporally stable phenomena (e.g., they "wax and wane in severity"; Fairburn et al., 1990, p. 406). Specifically, test–retest results indicated that Q-EDD diagnoses were quite stable over a 2-week period and less stable over a 1- to 3-month period. The 100% interscorer agreement across two studies indicates that scoring of the Q-EDD can be easily mastered.

In terms of incremental validity, when differentiating bulimics from all others, the Q-EDD and the BULIT-R performed quite similarly on all dimensions except positive predictive power, which was probably lower for the BULIT-R as a result of the inclusion of many EDNOS and anorexic individuals in this study. Hence, future researchers wanting only to distinguish bulimics from nonbulimics in a more general sample could use either the BULIT-R or the Q-EDD. Because of the low number of anorexics in Study 1, we could not directly compare the diagnostic accuracy of the EAT and the Q-EDD. However, the Q-EDD's high level of diagnostic accuracy for anorexia in Study 3, coupled with the EAT's basis in outdated diagnostic criteria and studies reporting it to have a high false-positive rate for diagnosing anorexia (e.g., Johnsone-Sabine et al., 1988; Meadows et al., 1986), leads to the tentative conclusion that the Q-EDD is a better measure of *DSM–IV* anorexia than is the EAT. Future research comparing the diagnostic accuracy of the two instruments with both clinical and nonclinical samples of anorexics should be conducted.

Perhaps our most important finding is the outstanding support we found for the criterion validity of the Q-EDD across both our clinical interview and clinician judgment studies. Specifically, for the differentiation of *DSM–IV* eating-disordered and nondisordered groups, the accuracy rates were 98% in Study 1 (clinical interview study) and 78% in Study 3 (clinician judgment study). For the distinction of eating-disordered, symptomatic, and asymptomatic groups, the accuracy rates were 90% in Study 1 and 78% in Study 3. Finally, for the differentiation of anorexia and bulimia, the accuracy rate was 100% in Study 3; this accuracy rate could not be examined in Study 1.

Both Study 1 and Study 3 revealed very good accuracy rates; however, differences across them are intriguing. With

respect to the differentiations of eating-disordered and non-disordered groups and eating-disordered, symptomatic, and asymptomatic groups, the clinical interview study yielded higher accuracy rates. Our hypothesis is that the interviewers were using more precise diagnostic criteria than were the clinicians, because the interviewers were trained to interview specifically in terms of a participant's meeting or not meeting *DSM–IV* criteria. On the other hand, the clinicians were given the criteria in written form but were not required to interview clients concerning the criteria; hence, clinicians' diagnostic judgments may have been based on previous interviews or on no diagnostic interviews at all. In support of this notion is a communication that Laurie B. Mintz had with one of the clinicians, in which this clinician indicated that she "guessed about if the client was subthreshold bulimic or bulimic" because she had not specifically gathered frequency information.

The Q-EDD appears to be quite effective at differentially diagnosing anorexia and bulimia. Across two studies, we found only one miss between anorexia and bulimia, and this individual was diagnosed by the Q-EDD as bulimic rather than anorexic only because her BMI of 17.7 barely exceeded the Q-EDD cutoff of 17.5. Of related interest is the fact that our clinical interview study included an individual with an almost identical Q-EDD profile (i.e., she met all criteria for binge-eating/purging-type anorexia, but her BMI was 17.7), and in this case the clinician indicated a diagnosis of bulimia. This difference in clinical judgments mirrors a controversy in the literature over the differentiation of bulimia and binge-eating/purging-type anorexia (see American Psychiatric Association, 1994; Fairburn & Garner, 1986). Future studies of the psychometric properties of the Q-EDD should include interviews with individuals who meet all of the Q-EDD scoring manual criteria for binge-eating/purging type anorexia except for missing the BMI cutoff by fractions; consistent interview diagnoses of anorexia might warrant raising the Q-EDD cutoff slightly (i.e., to 17.7 or 17.8).

Results concerning the differentiation of eating-disordered, symptomatic, and asymptomatic groups also provide interesting information to guide future research with the Q-EDD. Specifically, when misses between Q-EDD and clinical interview–clinician judgments occurred, the majority were between symptomatic and asymptomatic or between symptomatic and eating disordered. Across our two studies involving criterion validity, there were only three misses between asymptomatic and eating disordered. Test–retest changes followed a similar pattern: When changes in diagnoses occurred, the majority of movement was between asymptomatic and symptomatic and between symptomatic and eating disordered; few changes occurred between asymptomatic and eating disordered. Taken together, these results have two important implications. First, future researchers could use the Q-EDD to compare eating-disordered and asymptomatic women (i.e., eliminating symptomatic women), and the accuracy rates would be even higher. Indeed, if only eating-disordered and asymptomatic individuals had been compared, the accuracy rate for Study 1 would have been 99% and the accuracy rate in Study 3

would have been 94%. The second implication is that the Q-EDD may be a three-point eating disorder continuum measure.

Additional support for the Q-EDD as a three-point eating disorder continuum measure comes from our ANOVA of EAT scores by the Q-EDD groups of asymptomatic, symptomatic, and eating-disordered. Because the EAT has been conceptualized as a measure of general disordered eating in a nonclinical sample (Garner et al., 1982), our finding in Study 1 that EAT scores increased from the asymptomatic to symptomatic to eating-disordered groups supports the notion of the Q-EDD as a continuum measure. On the other hand, our finding in Study 2 of an increase in EAT scores from the asymptomatic group to both the symptomatic and eating-disordered groups, but not from the symptomatic group to the eating-disordered group, does not provide as strong support for this three-point continuum notion. Nevertheless, because both eating-disordered and symptomatic women can be conceptualized as engaging in disordered eating, this result is not surprising. An unpublished study by Zook (1995) found that scores on the Fear of Physical Unattractiveness subscale of Gillespie and Eisler's (1992) Feminine Gender Role Conflict Scale increased in a significant and linear fashion from Q-EDD asymptomatics to symptomatics to eating-disordered participants. In sum, there is preliminary evidence that the Q-EDD might be thought of as a continuum measure.

The idea that the Q-EDD might be conceptualized as a continuum measure is an exciting one. Although much has been written about the notion of an eating disorder continuum (Nylander, 1971; Rodin, Silberstein, & Striegel-Moore, 1985), there is a lack of systematic research on such a continuum, mainly as a result of the lack of an instrument with which to measure points along the continuum (Scarano & Kalodner-Martin, 1994). Also, although a few studies have examined such a continuum, the midpoint groups included in these studies have varied widely, making comparisons and generalizations across studies almost impossible. For example, in one study the groups were ideal eaters, dieters, presyndrome individuals, at-risk individuals, and bulimics (Hesse-Biber, 1989), whereas, in another, they were bulimics, overweight individuals, and sensible eaters (Nevo, 1985). In general, previous definitions of midpoint groups seemed to be guided more by researcher creativity than by the *DSM* classification system. The Q-EDD could prove to be a much needed *DSM–IV*-guided continuum measure that could be used across multiple studies. Future research should be undertaken to validate the Q-EDD as a three-point continuum measure; specifically, differences among the three groups with respect to a range of variables related to eating-disordered behaviors and correlates should be examined.

The results of our exploratory foci also have implications for future research with the Q-EDD. Specifically, one of our exploratory foci was to describe women with eating-disordered behaviors not captured by the *DSM–IV* (i.e., to describe symptomatic women). We thus generated descriptive labels based on eating disorder nomenclature and assessed these labels in terms of interscorer agreement and

agreement between Q-EDD scorers and clinical interview judgments. In both cases, a high level of agreement was found, indicating that our labels seem to capture identifiable groups. Interestingly, the misses between asymptomatic and symptomatic groups generally concerned different symptomatic subtypes (i.e., "restricting subthreshold nonbinging bulimia" and "chronic dieting") than did the misses between eating-disordered and symptomatic groups (i.e., "behavioral bulimia" and "nonnormal-weight nonbinging bulimia"). Also, misses between eating-disordered and symptomatic groups generally concerned what could be considered similar diagnoses such as binge-eating disorder and "subthreshold binge-eating disorder," bulimia and "behavioral bulimia," and nonbinging bulimia and "nonnormal-weight nonbinging bulimia." It appears that the lines between asymptomatic and symptomatic and between symptomatic and eating disordered may be fuzzy at times and that symptomatic subtypes may themselves exist on a continuum, with some being closer to asymptomatic and some being closer to eating disordered. Indeed, because the *DSM–IV* gives only examples of EDNOS categories and because these examples were strictly adhered to in generating decision rules for the Q-EDD, it is likely that the symptomatic category actually included some EDNOS examples that could be added to the list of *DSM–IV* examples.

On the basis of the results of this study, it appears that the symptomatic subtype of "nonnormal-weight nonbinging bulimia" should be thought of as an EDNOS. Indeed, the only difference between women in this category and those with the EDNOS diagnosis of nonbinging bulimia is weight; the severity of their behaviors is the same. "Behavioral bulimia" is probably another EDNOS; because the present study is the first to document the existence of this category, future research on it is clearly warranted. It is likely that this category has not previously been identified because, as noted earlier, few inventories have actually assessed the attitudinal component of the *DSM* criteria for bulimia (Fairburn et al., 1990). Future research could attempt to determine whether "behavioral bulimia" is a distinct EDNOS (i.e., an atypical form of bulimia) or whether "behavioral bulimics" are simply bulimics who are in denial. It should be noted that the results of this study support the former, because interviewers reported the existence of this disorder. Regardless of the findings of future research, one advantage of the Q-EDD is that research supporting the movement of specific symptomatic subgroups to the eating disorder category (i.e., as an EDNOS) would not result in changes in the Q-EDD. Rather, minor changes in the scoring manual are all that would be required. In short, we recommend that future researchers study symptomatic subgroups, such individuals having been largely ignored in previous research.

Future research is also needed to follow up on our second exploratory focus: the ability of the Q-EDD to differentially diagnose six eating disorder groups from one another. Initial results were promising, with a 90% accuracy rate in Study 1 and a 69% accuracy rate in Study 3. Accuracy rates increased to 94% in Study 1 and 86% in Study 3 when bulimia was combined with subthreshold bulimia. Study 3 accuracy rates increased even further (i.e., to 97%) when

anorexia was combined with menstruating anorexia. There is a great need in the eating disorder field for an instrument that can operationalize a full spectrum of eating disorders and make differential diagnoses (Fairburn et al., 1990; Williamson et al., 1995). The Q-EDD is the first questionnaire with the potential to be such an instrument, and future large-scale studies should be undertaken to further develop and validate it as such. Such future work should specifically attend to the Q-EDD's ability to distinguish between the major disorders (i.e., anorexia and bulimia) and their atypical and subthreshold forms, with particular attention paid to the differential diagnosis of bulimia and subthreshold bulimia; most within-eating-disorder criterion validity misses were between these two diagnoses.

Future research should examine the psychometric properties of the Q-EDD in other samples (e.g., men, community women, and high school students). Also, the Q-EDD is ideally suited to longitudinal research. Specifically, the Q-EDD could be administered repeatedly over a period of time, with each administration followed by interviews to confirm the diagnoses and patterns of change. More sophisticated research would involve examining the relations between diagnostic changes and risk and protective factors, such as stress and social support. Indeed, the Q-EDD may be an ideal questionnaire to help researchers adopt a "longitudinal and continuum measure approach. . .which has the promise of expanding our knowledge of the range of disturbed eating that occurs in the 'normal' population and increasing our understanding of the long-range course and causes of eating disorders" (Hesse-Biber, 1992, p. 389).

The Q-EDD can also be useful for counseling psychologists and others engaged in both clinical and outreach work. Because the Q-EDD yields both a diagnosis and frequency data for individual behaviors, it can be used to track progress in therapy. In addition, it could be administered at the start of therapy and then discussed in therapy; this would give the client and counselor a structure with which to discuss the often sensitive issue of types and frequency of behaviors. This would also aid in joint client–counselor goal setting concerning eating-disordered behaviors. The Q-EDD is quite straightforward and takes only 5 to 10 min to complete, making it an easy and noninvasive measure that could be used throughout therapy. The Q-EDD could also be used to target both at-risk and eating-disordered individuals so as to draw them into therapy and outreach programs. For example, the Q-EDD could be given to groups of women thought to be at risk for eating disorders, such as sorority women and certain groups of female athletes; if Q-EDDs revealed a high percentage of eating-disordered and symptomatic women, these women could be targeted for outreach efforts. The Q-EDD could also be given to large groups of 1st-year female high school and college students to determine the need for preventative efforts aimed at these individuals. It is possible that, whereas individuals with *DSM–IV* eating disorders would benefit from treatment, symptomatic individuals might benefit more from preventative psychoeducational interventions; of course, research is needed to confirm this. It is our

hope that the Q-EDD will be useful to both eating disorder clinicians and researchers.

References

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (Rev. 3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Button, E. J., & Whitehouse, A. (1981). Subclinical anorexia nervosa. *Psychological Medicine, 11*, 509–516.

Clarke, M. G., & Palmer, R. L. (1983). Eating attitudes and neurotic symptoms in university students. *British Journal of Psychiatry, 142*, 299–304.

Crews, T. M., & Sher, K. J. (1992). Using adapted short MASTs for assessing parental alcoholism: Reliability and validity. *Alcoholism: Clinical and Experimental Research, 16*, 576–584.

de Zwaan, M., Mitchell, J. E., Specker, S. M., Pyle, R. L., Mussell, M. P., & Seim, H. C. (1993). Diagnosing binge eating disorder: Level of agreement between self-report and expert-rating. *International Journal of Eating Disorders, 14*, 289–295.

Dykens, E. M., & Gerrard, M. (1986). Psychological profiles of purging bulimics, repeat dieters, and controls. *Journal of Consulting and Clinical Psychology, 42*, 431–439.

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, P. J. (1983). The epidemiology of bulimia nervosa: Two community studies. *International Journal of Eating Disorders, 2*, 61–67.

Fairburn, C. G., & Garner, D. M. (1986). The diagnosis of bulimia nervosa. *International Journal of Eating Disorders, 5*, 403–419.

Fairburn, C. G., Phil, M., & Beglin, S. J. (1990). Studies of the epidemiology of bulimia nervosa. *American Journal of Psychiatry, 147*, 401–408.

Feigner, J. P., Robin, E., Guze, S. B., Woodruff, R. A., Winokur, G., & Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry, 26*, 57–63.

First, M. B., Spitzer, R. L., Gibbon, R. L., Gibbon, M., & Williams, J. B. W. (1994). *Structured Clinical Interview for Axis I DSM-IV Disorders Patient Edition (SCID-I/P, Version 2.0)*. New York: Biometrics Research Department.

Freeman, C. P. L., & Henderson, M. (1988). The BITE: Indices of agreement. *British Journal of Psychiatry, 152*, 575–577.

Garner, D. M. (1991). *Eating Disorder Inventory-2: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Garner, D. M., & Garfinkel, P. E. (1979). The Eating Attitudes Test: An index of the symptoms of anorexia nervosa. *Psychological Medicine, 9*, 273–279.

Garner, D. M., Olmstead, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The Eating Attitudes Test: Psychometric features and clinical correlates. *Psychological Medicine, 12*, 871–878.

Garner, D. M., Olmstead, M. P., & Polivy, J. (1983). Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. *International Journal of Eating Disorders, 2*, 15–34.

Gillespie, B. L., & Eisler, R. M. (1992). Development of the Feminine Gender Roles Stress Scale. *Behavior Modification, 16*, 426–438.

Hawkins, R. C., & Clement, P. F. (1980). Development and construct validation of a self-report measure of binge eating tendencies. *Addictive Behaviors, 5*, 219–226.

Hesse-Biber, S. (1989). Eating patterns and disorders in a college population: Are college women's eating problems a new phenomenon? *Sex Roles, 20*, 71–90.

Hesse-Biber, S. (1992). Report on a panel longitudinal study of college women's eating patterns and eating disorders: Noncontinuum versus continuum measures. *Health Care for Women International, 13*, 375–391.

Johnsone-Sabine, E., Wood, K., & Patton, G. (1988). Abnormal eating attitudes in London schoolgirls—A prospective epidemiological study: Factors associated with abnormal response on screening questionnaires. *Psychological Medicine, 18*, 615–622.

Kaminski, P. L., & McNamara, K. (1996). A treatment for college women at risk for bulimia: A controlled evaluation. *Journal of Counseling and Development, 74*, 288–294.

Kaplan, R. M., & Saccuzzo, D. P. (1993). *Psychological testing* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Katzman, M. A., & Wolchik, S. A. (1984). Bulimia and binge eating in college women: A comparison of personality and behavioral characteristics. *Journal of Consulting and Clinical Psychology, 52*, 423–428.

King, M. B. (1986). Eating disorders in a general practice. *British Journal of Medicine, 293*, 1412–1414.

Kurtzman, F. D., Yager, J., Landsverk, J., Wiesmeier, E., & Bodurka, D. C. (1989). Eating disorders among selected female student populations at UCLA. *Journal of the American Dietetic Association, 89*, 45–53.

Lachenmeyer, J. R., & Muni-Brander, P. (1988). Eating disorders in a nonclinical adolescent population: Implications for treatment. *Adolescence, 23*, 303–312.

Leon, G. R., Fulkerson, J. A., Perry, C. L., & Dube, A. (1994). Family influences, school behaviors, and risk for the later development of an eating disorder. *Journal of Youth and Adolescence, 23*, 499–515.

Liebetrau, A. (1983). *Measures of association* (Sage University Paper Series on Quantitative Applications in the Social Sciences 07-032). Beverly Hills, CA: Sage.

Meadows, G. N., Palmer, R. C., Newball, E. U. M., & Kenrick, J. M. T. (1986). Eating attitudes and disorders in young women: A general practice based survey. *Psychological Medicine, 16*, 351–357.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficacy of psychometric signs, patterns, and cutting scores. *Psychological Bulletin, 52*, 194–216.

Mintz, L. B. (1989, August). Nonspecific eating disorders in two universities: Evidence for an epidemic. In L. Mintz (Chair), *Eating disorders: Clinically defined pathology as empirically normative behavior*. Symposium conducted at the 97th Annual Convention of the American Psychological Association, New Orleans, LA.

Mintz, L. B. & Betz, N. E. (1988). Prevalence and correlates of eating disordered behavior among undergraduate women. *Journal of Counseling Psychology, 35*, 463–471.

Moreno, A., & Thelen, M. H. (1993). Parental factors related to bulimia nervosa. *Addictive Behaviors, 18*, 681–689.

Nagelberg, D. B., Hale, S. L., & Ware, S. L. (1984). The assessment of symptoms and personality correlates in female college students. *Journal of Clinical Psychology, 40*, 440–445.

Nevo, S. (1985). Bulimic symptoms: Prevalence and ethnic differences among college women. *International Journal of Eating Disorders, 4*, 151–168.

Nylander, J. (1971). The feeling of being fat and dieting in a school

population: Epidemiological interview investigation. *Acta Sociomedica Scandinavica, 3,* 17–26.

O'Halloran, M. S. (1989, August). Prevalence and correlates of eating disorders among university women. In L. Mintz (Chair), *Eating disorders: Clinically defined pathology as empirically normative behavior.* Symposium conducted at the 97th Annual Convention of the American Psychological Association, New Orleans, LA.

Ousley, L. B. (1986). *Differences among bulimic subgroups, binge-eaters, and normal dieters in a female college population.* Unpublished doctoral dissertation, University of Florida, Gainesville.

Patton, G. C., & King, M. B. (1991). Epidemiological study of eating disorders: Time for a change of emphasis [Editorial]. *Psychological Medicine, 21,* 287–291.

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.

Pyle, R. L., Halvorson, P. A., Neuman, P. A., & Mitchell, J. E. (1986). The increasing prevalence of bulimia in freshman college students. *International Journal of Eating Disorders, 5,* 631–647.

Rodin, J., Silberstein, L. R., & Striegel-Moore, R. H. (1985). Women and weight: A normative discontent. In T. B. Sonderegger (Ed.), *Nebraska Symposium on Motivation: Vol. 32. Psychology and gender* (pp. 267–307). Lincoln: University of Nebraska Press.

Scarano, G. M. (1991). *The eating disorder continuum: Relationship to food/body and psychological issues.* Unpublished master's thesis, University of Akron, Akron, OH.

Scarano, G. M., & Kalodner-Martin, C. R. (1994). A description of the continuum of eating disorders: Implications for intervention and research. *Journal of Counseling and Development, 72,* 356–361.

Schotte, D. E., & Stunkard, A. J. (1987). Bulimia vs. bulimic behaviors on a college campus. *Journal of the American Medical Association, 258,* 1213–1215.

Shefer, T. (1987). Abnormal eating attitudes and behaviors among women students. *South African Medical Journal, 72,* 419–421.

Skre, I., Onstad, S., Torgersen, S., & Kringlen, E. (1991). High interrater reliability for the Structured Clinical Interview for DSM-III-R Axis (SCID-I). *Acta Psychiatrica Scandinavica, 84,* 167–173.

Smithies, C. S. (1989, August). Prevalence and correlates of eating disorders among synchronized swimming athletes. In L. Mintz (Chair), *Eating disorders: Clinically defined pathology as empirically normative behavior.* Symposium conducted at the 97th Annual Convention of the American Psychological Association, New Orleans, LA.

Spitzer, R. L., Delvin, M., Walsh, B. T., Hasin, D., Wing, R., Marcus, M., Stunkard, A., Wadden, T., Yanovski, S., Agras, S., Mitchell, J., & Nonas, C. (1992). Binge eating disorder: A multisite field trial of diagnostic criteria. *International Journal of Eating Disorders, 11,* 191–203.

Spitzer, R. L., Yanovski, S., Wadden, T., Wing, R., Marcus, M., Stunkard, A., Delvin, M., Mitchell, J., Hasin, D., & Horne, R. L. (1993). Binge eating disorder: Its further validation in a multisite study. *International Journal of Eating Disorders, 13,* 137–153.

Striegel-Moore, R. H., Silberstein, L. R., Frensch, P., & Rodin, J. (1989). A prospective study of disordered eating among college students. *International Journal of Eating Disorders, 8,* 499–509.

Szmukler, G. I. (1982). Weight and food preoccupation in a population of English schoolgirls. In D. E. Redfom (Ed.), *Understanding anorexia nervosa and bulimia* (pp. 21–28). Columbus, OH: Ross Laboratories.

Thelen, M. H., Farmer, J., Wonderlich, S., & Smith, M. (1991). A revision of the Bulimia Test: The BULIT-R. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3,* 119–124.

Thelen, M. H., Mintz, L. B., & Vander Wal, J. S. (1996). The Bulimia Test-Revised: Validation with DSM-IV criteria for bulimia nervosa. *Psychological Assessment, 8,* 219–221.

Walsh, W. B., & Betz, N. E. (1985). *Tests and assessment.* Englewood Cliffs, NJ: Prentice Hall.

Whitehouse, A. M., & Button, E. J. (1988). The prevalence of eating disorders in a U. K. college population: A reclassification of an earlier study. *International Journal of Eating Disorders, 7,* 393–397.

Widiger, T. A., Hurt, S. W., Frances, A., Clarkin, J. F., & Gilmore, M. (1984). Diagnostic efficiency in DSM-III. *Archives of General Psychiatry, 41,* 1005–1012.

Williams, R. L., Schaefer, C. A., Shisslak, C. M., Gronwaldt, V. H., & Comerci, G. D. (1986). Eating attitudes and behaviors in adolescent women: Discrimination of normals, dieters, and suspected bulimics using the Eating Attitudes Test and Eating Disorder Inventory. *International Journal of Eating Disorders, 5,* 879–894.

Williamson, D. A. (1990). *Assessment of eating disorders: Obesity, anorexia, and bulimia-nervosa.* New York: Pergamon Press.

Williamson, D. A., Anderson, D., Jackman, L. P., & Jackson, S. R. (1995). Assessment of eating disordered thoughts, feelings, and behaviors. In D. Allison (Ed.), *Handbook of assessment methods for eating behaviors and weight-related problems* (pp. 303–346). Thousand Oaks, CA: Sage.

Yager, J., Landsverk, J., & Edelstein, C. K. (1987). A 20-month follow-up study of 628 women with eating disorders, I: Course and severity. *American Journal of Psychiatry, 144,* 1172–1177.

Zook, C. E. (1995). *Disordered eating behaviors and perceived gender role stress in college women.* Unpublished master's thesis, University of Missouri—Columbia.

Zuckerman, D. M., Colby, A., Ware, N. C., & Lazerson, J. S. (1986). The prevalence of bulimia among college women. *American Journal of Public Health, 76,* 1135–1137.