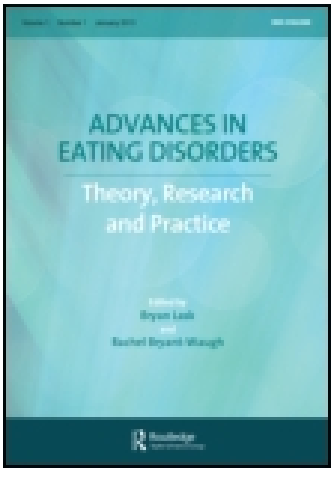

Advances in Eating Disorders: Theory, Research and Practice
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/reat20

The Eating Disorder Examination: a semi-structured interview for the assessment of the specific psychopathology of eating disorders

Jennifer J. Thomas[ab], Christina A. Roberto[c] & Kelly C. Berg[d]
[a] Eating Disorders Clinical and Research Program, Massachusetts General Hospital, Boston, MA, USA
[b] Department of Psychiatry, Harvard Medical School, Boston, MA, USA
[c] Department of Social and Behavioral Sciences, Harvard School of Public Health, Boston, MA, USA
[d] Department of Psychiatry, University of Minnesota, Minneapolis, MN, USA
Published online: 07 Nov 2013.

To cite this article: Jennifer J. Thomas, Christina A. Roberto & Kelly C. Berg (2014) The Eating Disorder Examination: a semi-structured interview for the assessment of the specific psychopathology of eating disorders, Advances in Eating Disorders: Theory, Research and Practice, 2:2, 190-203, DOI: 10.1080/21662630.2013.840119

To link to this article: http://dx.doi.org/10.1080/21662630.2013.840119

Advances in Eating Disorders: Theory, Research and Practice, 2014
Vol. 2, No. 2, 190–203, http://dx.doi.org/10.1080/21662630.2013.840119

Routledge
Taylor & Francis Group

The Eating Disorder Examination: a semi-structured interview for the assessment of the specific psychopathology of eating disorders

Jennifer J. Thomas[a,b]*, Christina A. Roberto[c] and Kelly C. Berg[d]

[a]*Eating Disorders Clinical and Research Program, Massachusetts General Hospital, Boston, MA, USA;* [b]*Department of Psychiatry, Harvard Medical School, Boston, MA, USA;* [c]*Department of Social and Behavioral Sciences, Harvard School of Public Health, Boston, MA, USA;* [d]*Department of Psychiatry, University of Minnesota, Minneapolis, MN, USA*

(*Received 3 July 2013; accepted 28 August 2013*)

The year 2013 marks not only the twenty-fifth anniversary of the Eating Disorder Examination (EDE) but also the publication of DSM-5. This fortuitous confluence presents the perfect opportunity to reflect on the current status of eating disorder assessment. As the first structured interview to evaluate the unique psychopathology of anorexia nervosa and bulimia nervosa, the EDE revolutionised clinical understanding and research methodology in the eating disorder field. However, like any measure designed to capture illusive and evolving psychological constructs, the EDE has theoretical, logistical, and functional limitations. We review each limitation in turn and make proposals for the next generation of eating disorder assessment – from substantially streamlining the EDE to the potentially heretical consideration of alternative measures.

Keywords: Eating Disorder Examination; DSM-5; anorexia nervosa; bulimia nervosa; other specified feeding or eating disorder; assessment

The initial paper on the development of the Eating Disorder Examination (EDE; Cooper & Fairburn, 1987) has arguably been one of the most influential in our field. With nearly 700 citations in the past 25 years according to Google Scholar, the EDE is often referred to as 'the gold standard' of eating disorder (ED) assessment (e.g. Stice, Telch, & Rizvi, 2000). The purpose of the Classics Revisited section is to evaluate whether seminal papers such as this one have withstood the test of time. The fortuitous 2013 confluence of the EDE's twenty-fifth anniversary and the publication of *DSM-5* present the perfect opportunity to reflect on the current status of ED measurement. Collectively, the three of us have conducted more than 600 EDEs (~200 JJT; ~250 CAR; and ~200 KCB), and published extensively on the diagnosis and assessment (e.g. Berg, Peterson, Frazier, & Crow, 2011, 2012; Roberto, Grilo, Masheb, & White, 2010; Roberto, Steinglass, Mayer, Attia, & Walsh, 2008; Thomas, Roberto, & Brownell, 2009; Thomas, Vartanian, & Brownell, 2009), putting us in a unique position to provide this critique. The goal of this paper is therefore to pay homage to the EDE and its revolutionary impact on the field; identify theoretical, logistical, and functional limitations of the measure 25 years after its initial publication; and provide recommendations for the next generation of ED assessment.

*Corresponding author. Email: jjthomas@partners.org

Homage to the EDE

The EDE was the first structured interview to assess the specific psychopathology associated with EDs. Before the EDE, the only structured assessments of ED symptoms were self-report questionnaires (e.g. Eating Attitudes Test, Garner & Garfinkel, 1979; Eating Disorder Inventory, Garner, Olmstead, & Polivy, 1983). Although there are obvious advantages to paper-and-pencil inventories (e.g. time and cost efficiency and perceived anonymity), they have limitations. Specifically, in their original paper, Cooper and Fairburn (1987) pointed out that 'it may be doubted whether … questions are interpreted in a standard way by all respondents' (p. 3). In response to this concern, the pioneering EDE authors provided specific parameters to define complex constructs such as 'binge eating'. They also designed the EDE as an *investigator-based* interview, which allows the assessor to ask additional, clarifying questions and requires the assessor to have specialised training in the constructs being assessed. A second concern regarding self-report questionnaires is temporal sensitivity: Cooper and Fairburn (1987) noted that, 'Clearly self-report questionnaires that simply enquire about the frequency of binge eating will not detect subtle changes [in the amount of food consumed and the extent to which loss of control is experienced]' (p. 4). As a result, the EDE not only collects quantitative data (e.g. binge frequency), but also collects qualitative data (e.g. amounts eaten and perceived loss of control), as well as detailed information on the cognitive symptoms of EDs such as body dissatisfaction, over-evaluation of shape and weight, dietary restraint, and concerns about eating in front of others.

The EDE also transformed the ED assessment process by beginning the interview with an orientation to the time frame. A calendar is provided and respondents are asked to describe any events during that month that would help them to remember the time period (e.g. days off of work/school, holidays, and major life events). The objective of the orientation was to minimise recall bias (e.g. Stone & Shiffman, 1994; Teasdale & Fogarty, 1979), as prior research in alcohol dependence had found that orienting participants to the time period relevant to the assessment was associated with higher test-retest reliability (e.g. Sobell, Maisto, Sobell, & Cooper, 1979; Sobell, Sobell, Klajner, Pavan, & Basian, 1986). Another key advancement of the EDE was that it could be used both dimensionally (e.g. EDE subscales and behaviour frequency scores) and categorically (e.g. to derive ED diagnoses). This characteristic enhances the utility of the EDE and reduces the need for multiple ED assessments, which contribute to respondent burden. Furthermore, the EDE eliminated skip logic, which substantially limits the quality and utility of the data collected (e.g. Swanson, Crow, Le Grange, Swendsen, & Merikangas, 2011). The elimination of skip logic is particularly important for the clinicians and researchers interested in subthreshold ED presentations because skip logic is often implemented when criteria are not met for full-threshold ED. Finally, although additional data are needed to broaden the generalisability of the findings, there is a body of research that supports aspects of the reliability and validity of the EDE as an assessment of ED psychopathology (for review, see Berg, Peterson et al., 2012).

The EDE has not only impacted assessment, but it has also revolutionised the ED field as a whole. For example, the operationalisation of key diagnostic constructs likely influenced the *DSM-IV* and *DSM-5* ED criteria, which more closely approximate the EDE descriptions of binge eating and over-evaluation of shape and weight than those described in *DSM-III-R*. Similarly, the EDE identified values that patients prioritise (e.g. 'to exert self-control is a sign of strength and discipline', Cooper & Fairburn, 1987, p. 3) but might not spontaneously articulate (i.e. 'patients are not necessarily aware of their presence or influence since they are so much a part of their conceptual scheme that they are usually unable to identify and question them', Cooper & Fairburn, 1987, p. 3). Additionally, because the EDE quickly became the 'gold-standard' ED assessment, it was used consistently – particularly in treatment outcome studies. This persistent use of the EDE provided the opportunity to compare outcomes across studies with

enhanced accuracy, which has been critically important to advancing the field. This is to be contrasted with outcome studies in other fields (e.g. bariatric surgery), in which outcome is measured inconsistently (e.g. change in body mass index, change in percentage of expected weight loss, and change in raw weight), making it very difficult to make meaningful comparisons across studies. Finally, the authors of the EDE elected to place the EDE in the public domain and allow anyone to use it free of charge. This intellectual generosity has been incredibly beneficial, as it effectively democratised structured ED assessment. Additionally, the decision to put the EDE in the public domain set a precedent in the field and a number of subsequent ED assessments were also offered free of charge (e.g. Eating Disorder Diagnostic Scale, Stice et al., 2000 and Structured Interview for Anorexic and Bulimic Symptoms, Fichter et al., 1991).

Limitations of the EDE 25 years later

Although the EDE has undeniably transformed our understanding of EDs, like any measure aiming to assess illusive and evolving psychological constructs, it has important limitations. These can be divided into theoretical, logistical, and functional.

Theoretical limitations

Perhaps the most important limitations are theoretical. These disadvantages primarily reflect the historical context in which the measure was developed – i.e. as a measure of cognitive behavioural therapy (CBT) response among women with bulimia nervosa (BN). Such limitations thus include: (1) a bias towards assessing the psychopathology of BN; (2) a partiality towards concepts most relevant to CBT; (3) an emphasis on stereotypically feminine manifestations of body image disturbance; and (4) a lack of evidence for the proposed factor structure.

First, the EDE was originally developed to assess the psychopathology of anorexia nervosa (AN) and BN. To create the initial item pool, 'the literature in English on anorexia nervosa and bulimia nervosa was studied to obtain a comprehensive description of the specific psychopathology of the two disorders' (Cooper & Fairburn, 1987, p. 4). Next, 'a series of lengthy unstructured interviews was conducted with patients with anorexia nervosa and bulimia nervosa with the aim of eliciting detailed descriptions of various aspects of their behavior and attitudes that might form the basis for further items' (Cooper & Fairburn, 1987, p. 4). Importantly, many of the items (e.g. the concept of attempting to restrict, whether or not one succeeds) are much more relevant to BN than to AN. Indeed, a robust literature indicates that individuals with AN typically score *lower* on the EDE in comparison to those with BN (e.g. Cooper, Cooper, & Fairburn, 1989). Although the authors have been careful to emphasise that this does *not* mean that one is 'more ill' than the other, they just have different EDE profiles (O'Connor, 2009), the frequent application of clinical cut points (i.e. subscale scores of four or higher indicating likely psychopathology, e.g. Fairburn & Cooper, 1993; Mitchell et al., 2011) suggest that these differential profiles are a drawback of the measure. Indeed, since scoring within one standard deviation of EDE community norms is now being used as a criterion for remission in transdiagnostic treatment studies (e.g. Fairburn et al., 2009), it seems somewhat problematic that, in two recent trials of CBT for AN, 42% of adolescents (Dalle Grave, Calugi, Doll, & Fairburn, 2013) and 33% of adults (Fairburn et al., 2013) already scored in the normative range before treatment even took place. In sum, although it made sense for the EDE to focus primarily on AN and BN three decades ago when *DSM-III-R* contained so few ED diagnoses (i.e. AN, BN, and eating disorder not otherwise specified [EDNOS]), this focus may be overly narrow now that *DSM-5* has recognised a greater diversity of presentations including AN, BN, binge eating

disorder (BED), and avoidant/restrictive food intake disorder (ARFID); and also further differentiated EDNOS (now called as other specified feeding or eating disorder) into five named subtypes (atypical AN, subthreshold BN, subthreshold BED, purging disorder, and night eating syndrome). Subsequent versions of the EDE could certainly add *DSM-5* constructs, but our understanding of diagnostic heterogeneity has advanced so much in the past three decades that it may no longer be realistic to think that we can capture all meaningful ED symptoms in a single interview.

Second, the EDE has a clear bias towards constructs that are relevant to CBT. For example, individual items such as reaction to prescribed weekly weighing specifically reflect key CBT interventions. Moreover, because the CBT model identifies failed adherence to rigid dietary rules as the primary trigger for binge eating, many items are designed to measure dietary *restraint* (i.e. initial attempts to cut back and regardless of ultimate success) rather than dietary *restriction* (i.e. successful reduction in calorie intake leading to a negative energy balance). Subsequent research has demonstrated that self-report measures of dietary restraint are often uncorrelated with laboratory measures of dietary restriction (Stice, Fisher, & Lowe, 2004), suggesting that an important facet of ED psychopathology (i.e. restriction and a core characteristic of AN) is explicitly de-emphasised in the EDE. Another consequence of the EDE's CBT focus is that, necessarily, cognitive and behavioural features take precedence. A growing body of evidence from ecological momentary assessment paradigms suggest that increases in negative affect throughout the day are associated with a greater likelihood of restricting (Engel et al., 2013), binge eating (Berg, Crosby et al., 2013; Haedt-Matt & Keel, 2011; Smyth et al., 2007), and purging (Berg, Crosby et al., 2013). Several newer treatment packages such as interactive cognitive affective therapy (Wonderlich et al., 2009) and emotion acceptance behaviour therapy (Wildes & Marcus, 2011) – and even Fairburn's own 'enhanced' version of CBT (Fairburn, 2008) – recognise the importance of 'emotional eating' (i.e. eating in response to changes in affect), a concept not explicitly assessed on the EDE.

Third (and surely this critique could be levied against most contemporary measures of eating pathology), the EDE is much more relevant to stereotypically feminine manifestations of body image disturbance. Items such as 'flat stomach' and 'desire to lose weight' might feel less germane to boys and men, who favour muscularity (e.g. six pack abs) over thinness. In one study, adolescent males enrolled in AN treatment trials earned lower scores than their female counterparts on all EDE subscales, despite being matched on a primary marker of clinical severity (i.e. per cent expected body weight) (Darcy et al., 2012).

Lastly, the proposed EDE factor structure (i.e. comprising the four domains of Restraint, Eating Concern, Shape Concern, and Weight Concern) has never been empirically validated (Berg, Peterson et al., 2012). Incredibly, investigators continue to dutifully report scores on all four subscales in studies as diverse as treatment trials, neuroimaging, and genetics. This flies in the face of psychometric findings, which indicate that Pearson *r* correlations between subscales (Cooper, Cooper, & Fairburn, 1989) exhibit medium to large effect sizes. In particular, the Shape Concern and Weight Concern subscales have never been extracted as separate components in a factor analysis (Berg, Peterson et al., 2012), likely due to their considerable conceptual overlap. Not only do many patients view shape and weight as interchangeable, but the two scales also contain similarly worded items that may contribute to method variance (e.g. preoccupation, over-evaluation, and dissatisfaction items), and even share an item that counts towards both subscales (preoccupation with shape and weight). Oddly, one item that specifically mentions weight rather than shape (i.e. fear of weight gain) counts towards the Shape Concern (but not the Weight Concern) subscale. One study indicated that a one-factor model provided the best fit, leading investigators to conclude that individual EDE subscales should be eschewed in favour of the global score (Byrne, Allen, Lampard, Dove, & Fursland, 2010). Indeed, the use of the

global score in lieu of subscale scores is further supported by psychometric studies of the Eating Disorder Examination-Questionnaire (EDE-Q), which have also failed to replicate the proposed four-factor structure and support the interpretation of the global score as a unitary construct (Aardoom, Dingemans, Landt, & Van Furth, 2012; Becker et al., 2010; Friborg, Reas, Rosenvinge, & Ro, in press).

Logistical limitations

The next series of EDE critiques reflect logistical problems. The primary concerns involve the interview's burden with respect to the (1) clinical value of some items, (2) item redundancy, (3) overly complex questions, (4) rapport-rupturing items, and (5) accessibility of specialised training.

When designing an assessment tool, one must balance the goal of obtaining key clinical information with minimising assessor/respondent burden. The EDE initially started out as a 62-item assessment (Cooper & Fairburn, 1987). Understandably, as the knowledge of the clinical features of EDs grew over time, so too did the EDE. Now in its sixteenth version, it has over 100 items and is 44 pages long. The original description of the EDE indicated that, 'the interview takes between 30 minutes and 1 hour to complete' (Cooper & Fairburn, 1987, p. 6). In our experience, administering the EDE v. 16 almost always takes a full hour to complete as intended and can often take longer. Although the rich and diverse array of questions are clinically interesting, it is important to consider whether each question provides enough essential information to justify the assessment burden on respondents and assessors. What is considered as 'essential' information will vary based on the role of the assessor, since researchers and clinicians can be interested in different types of information and levels of detail.

In its current format, we believe that there are several ways in which the EDE puts an undue burden on assessors and respondents alike. First, the clinical utility of some items remains unclear, particularly newer or infrequently reported items such as 'How concerned have you been about the composition of your body?'; 'Over the past month have you felt that any particular part of your body is too fat?'; and 'Over the past four weeks what amount of weight gain, over a period of one week, would have definitely upset you?' Although the items might be capturing clinically relevant information, the novelty and importance of that information, relative to clinical data ascertained from other EDE items/subscales, is unclear.

The EDE also increases its time burden by requiring the assessment of highly redundant information within items. For example, both days and episodes for objective binges, subjective binges, and overeating are assessed, despite data indicating that these scores are correlated in the .90 range (Berg, unpublished data) as well as findings suggesting they can be used interchangeably (Berg, 2010). For diagnostic purposes, the *DSM-5* now only focuses on episodes, suggesting that the assessment of days (which was relevant to *DSM-IV* BED only) could be removed. Another issue is between item redundancies. For example, respondents are first asked whether they avoid certain foods altogether. This is followed by a question asking whether they follow any definite dietary rules. Individuals often report rules about avoiding certain foods, which can lead to the same information getting coded twice.

In addition, a lot of time is spent on the 'importance of weight, shape, and strict control over eating' items. These items are identified as diagnostic and are meant to capture whether body weight and/or shape unduly influence self-evaluation. In our experience, we have found that many individuals struggle to understand these items. Indeed, the EDE instructions acknowledge this by prompting the assessor to explain to the patient that they are about to be asked 'a rather complex question'. This difficult and time-consuming item is administered for weight and then repeated for shape and for 'strict control over eating', adding to the length and complexity of

the interview, without a strong empirical rationale for separating weight, shape, and control. Another concern is that the item asks assessors to list aspects that are important to one's self-worth. In our experience, this listing exercise can generate hesitancy to place weight or shape at the top of the list, out of fear of shame or embarrassment during the interview. These concerns suggest that it might be worth trying to improve the item's clarity or determine whether a clinician can make a diagnostic call about this construct based on other clinical details ascertained during the interview.

Furthermore, there are some items and prompts that have the potential to disrupt rapport. For example, questions about an individual's level of dissatisfaction with his/her body weight or shape include the suggested prompt: 'Could you have felt worse?' Although we see the value in the question's ability to distinguish between levels of severity, we have found that the wording can sometimes come across as insensitive to patients, even when taking care to ask it with clinical sensitivity. In addition, the placement of other questions too late in the interview can damage rapport. For example, patients are asked at the end whether they have been 'trying to lose weight'. This kind of information typically emerges much earlier in the interview, giving an impression of redundancy (at best) or a lack of attentiveness (at worst). Similar issues apply to the 'feeling fat' item appearing towards the end of the interview, and the query 'Why have you been dissatisfied with your weight?' after an affirmative response to weight dissatisfaction. Some items related to weight and shape have the potential to be especially disruptive to rapport when interviewing obese patients with BED or individuals seeking bariatric surgery. For example, items that inquire about feelings of fatness, regional fatness, or body composition are often met with confusion or discomfort by patients who are overweight/obese. Asking these kinds of questions of patients who meet objective thresholds for overweight/obesity can make the interviewer appear insensitive or out of touch.

Lastly, although the interview itself is available free of charge, the authors note that 'if … the EDE is going to be used for research purposes, training is essential' (Fairburn, Cooper, & O'Connor, 2008, p. 1). While we certainly agree that specialised training is vital for ensuring competent administration, such training may not be accessible to all researchers desiring to use the measure. Specifically, the training is costly (approximately $300 US to take the course at Oxford), time-consuming (1–2 full days), and not widely available outside of major hospitals and universities.

Functional limitations

In addition to the theoretical and logistical problems described above, there are also issues that affect the usefulness of the EDE in both research and clinical settings. These issues largely fall into the following three categories: (1) limitations of the EDE as a diagnostic instrument, (2) limitations of individual EDE items, and (3) limitations of the anchors used with the EDE Likert scales.

The EDE was originally designed as a dimensional measure of ED pathology and the authors stressed that it was 'not intended to be used as a diagnostic instrument' (Cooper & Fairburn, 1987, p. 7). However, the authors have since indicated that the EDE can, in fact, 'be used to generate operationally defined eating disorder diagnoses' (Fairburn & Cooper, 1993, p. 319). To facilitate its use as a diagnostic instrument, the authors have published recommendations for diagnostic algorithms (Fairburn & Cooper, 1993) and have highlighted *diagnostic items* within the interview schedule itself (Fairburn & Cooper, 1993; Fairburn et al., 2008). Although there are presumed advantages to using the EDE as a diagnostic measure (e.g. enhanced inter-rater reliability compared to unstructured clinical interviews, more accurate estimates of binge eating frequency as compared to self-report questionnaires), there are also important limitations to using the EDE

diagnostically. First, and perhaps most importantly, the EDE diagnostic algorithms are not entirely representative of *DSM* criteria. For example, the 'Dietary Restriction Outside Bulimic Episodes' item is used to measure the average amount of food a respondent consumes on a typical day, not taking into account food consumed during episodes of binge eating. The EDE purports that this item represents an index of fasting, a non-purging compensatory behaviour that is included in the *DSM-IV* and *DSM-5* criteria for BN (Fairburn & Cooper, 1993, p. 321). However, one can imagine a situation in which every day a respondent binges multiple times per day and does not consume anything in between binge episodes. In this scenario, the correct code for this behaviour would be a '2', which the EDE describes as 'purposeful "fasting"' and meets the EDE criteria for a diagnosis of BN (Fairburn & Cooper, 1993, p. 321). It is unlikely, however, that this behaviour reflects what is meant by 'fasting' in the DSM. The inconsistencies between the EDE algorithms and the *DSM* criteria may be due, in part, to the fact that Fairburn and Cooper (1993) define the algorithms as 'the rules currently in use in Oxford for generating *DSM-III-R* and *DSM-IV* diagnoses' (Fairburn & Cooper, 1993, p. 319) and then go on to provide specific examples of how the Oxford criteria differ from the *DSM* criteria (p. 321). Alternative diagnostic algorithms that more closely resemble the *DSM* criteria have been published (Berg, Stiles-Shields et al., 2012) for both *DSM-IV* and *DSM-5*; however, fully replicating the *DSM* criteria is impossible because the EDE is not fully comprehensive of *DSM* constructs (e.g. lack of recognition of seriousness of low body weight in AN).

A second limitation of the EDE's diagnostic algorithms is that the EDE was 'designed to assess the present state of patients and as such is only concerned with the preceding 4 weeks. This time period was chosen since it takes account of week-to-week fluctuations in symptoms while not being so long as to raise doubts about the accuracy of patients' recall of symptoms (Cooper & Fairburn, 1987, p. 7). The ability of respondents to recall symptom fluctuations is especially relevant to the assessment of binge eating because specific frequencies of these behaviours are required for diagnoses of BN and BED. Although research suggests that the 7- to 28-day frequency of binge eating reported on the EDE and daily records are significantly correlated (Farchaus Stein & Corte, 2003; Loeb, Pike, Walsh, & Wilson, 1994; Rosen, Vara, Wendt, & Leitenberg, 1990), respondents also appear to report significantly more binge eating episodes when assessed with the EDE vs. daily records (Berg, Peterson, & Crow, 2009; Farchaus Stein & Corte, 2003). Furthermore, diagnoses of BN and BED require binge eating frequency estimates for the past three months (American Psychiatric Association [APA], 2013), yet research suggests that the relationship between binge eating frequency scores on the EDE and daily records weakens as the assessment period becomes more distal (Berg, 2010). In summary, the significant correlations between binge eating frequency scores on the EDE and daily records provide evidence for the validity of the EDE as a dimensional measure of binge eating frequency. Unfortunately, these data may give the illusion that using the EDE enhances the accuracy of retrospective binge eating assessments: How many of us have data-sets with an official-looking column of numbers purporting to represent 'binge frequency in month 6'?! In summary, significant differences between mean binge eating frequency scores on the EDE and daily records and the impact these differences can have on diagnoses (Berg, 2010; Berg et al., 2009) call into question the validity of the EDE as a diagnostic instrument.

A third limitation to using the EDE diagnostically is the fact that an 'EDE month' is only 28 days long. Although this time frame does apply to three out of four Februaries, the other calendar months are all longer than 28 days, ranging from 29 to 31 days each. As stated earlier, the *DSM* criteria for both BN and BED require a specific frequency of behaviours over the past three months (APA, 2013), and although it is not specified, it is assumed that the *DSM* operates using calendar months. It is easy to imagine a scenario in which an individual meets the frequency criteria for BN or BED using calendar months, in which three months is equivalent to 90–92 days,

but not EDE months, in which three months is equivalent to 84 days. In other words, using the EDE to assess behaviour frequencies for the past three months may, in essence, short-change respondents in terms of the amount of time they have to meet the criteria for BN and BED. As a result, they may appear less ill than they would if they were assessed using calendar months. Thus, even though the difference between the calendar months and the EDE months is relatively small (6–8 days), it is a critically important difference that may impact diagnostic status and subsequently, treatment accessibility, insurance coverage, research eligibility, etc.

A final limitation of the EDE diagnostic algorithms concerns the '*DSM-IV* "Binge Eating Disorder" Module'. This section was not originally included in the EDE (Fairburn & Cooper, 1993) and is the only section of the EDE that is optional. Perhaps as a result, one is left with the impression that the section was simply tacked on out of necessity rather than being thoughtfully and seamlessly incorporated into the existing interview.

As stated earlier, the EDE includes a substantial number of items that are not included in the EDE subscales and are not used in any diagnostic algorithms. As such, these items are only potentially useful as single-item measures of specific constructs. Although single-item measures are not generally considered robust, individual items from the EDE are often used as indicators in latent structure analyses (e.g. Eddy et al., 2010; Striegel-Moore et al., 2005) or to derive diagnostic criteria (e.g. Crow, Agras, Halmi, Mitchell, & Kraemer, 2002; Mitchell et al., 2011). Although the inter-rater reliability of scores on some of the individual EDE items has been examined (Cooper & Fairburn, 1987), there is no empirical evidence for the validity of any individual EDE item as a measure of ED pathology (Berg, Peterson et al., 2012). One such example is the 'Picking (Nibbling)' item which measures a specific type of eating that is unplanned (both in terms of the occurrence of the eating as well as the amount consumed), repetitious, and not characterised by feelings of loss of control (Fairburn et al., 2008; p. 11). In a sample of 217 adults with concurrent BED and obesity, scores on 'Picking (Nibbling)' were not significantly correlated with any measure of ED pathology, including BMI and binge frequency (Masheb, Roberto, & White, in press). Similarly, Reas, Wisting, Kapstad, and Lask (2012) found that 91% of a normative sample of young adult women reported engaging in nibbling, and that nibbling behaviour was not associated with BMI, binge frequency, compensatory behaviours, meal or snack patterns, or EDE subscales. Other specific examples of items that have not been validated are the 'control items' which were recently added to the Restraint subscale; 'Maintaining Strict Control over Eating'; 'Sensitivity to Weight Gain'; 'Regional Fatness'; 'Vigilance about Shape'; and 'Body Composition'.

The usefulness of other individual EDE items is questionable due to the operationalisation of the constructs themselves. For example, the EDE guidelines specifically indicate that 'Drinks do not count as food' (Fairburn et al., 2008). This instruction means that attempts to restrict fluids (e.g. sodas and juices) do not count towards ratings of the Restraint subscale items (e.g. 'Restraint over Eating', 'Food Avoidance', and 'Dietary Rules') despite data indicating that individuals with ED often restrict or attempt to restrict fluids (Lowinger, Griffiths, Beumont, Scicluna, & Touyz, 1999). Nor would the restriction of calorie-laden beverages as exemplified in the anorexia memoir named after the dietary maxim so popular among individuals with ED: *Never Drink Calories* (Otto, 2011). It also means that fluid consumption is not considered when determining whether an amount of food is large enough to be considered as an objective bulimic episode, even though evidence indicates that the consumption of sugary beverages is associated with obesity (Ludwig, Peterson, & Gortmaker, 2001).

Lastly, there are several issues regarding the way in which EDE items are rated that impact the utility of the EDE. First, the EDE instructs the respondent to code both the frequency (e.g. 'Dietary Rules' and 'Feeling Fat') and severity (e.g. 'Dissatisfaction with Weight' and 'Discomfort Seeing Body') using a Likert scale with unequal spacing[1] and an absolute zero point (Fairburn et al., 2008; O'Connor, 2009). This creates a situation in which the level of measurement is unclear; the unequal

spacing suggests that the EDE uses an ordinal scale, whereas the absolute zero point suggests a ratio scale. This is problematic because different analytic techniques are required for ordinal (e.g. non-parametric statistics) vs. interval-ratio (e.g. parametric statistics) scales (Tabachnick & Fidell, 2001). Second, the sixteenth edition of the EDE changed such that 'For severity items, ratings now based on mode over the previous 28 days rather than mean' (Fairburn et al., 2008, p. 308). Although this switch simplifies the rating process (and may, in turn, increase inter-rater reliability), it also results in a loss of information because the rating only represents what happened on 'most' days and completely ignores anything that happened on the 'other' days.

Finally, there are some items for which clinically relevant data are given an alternative score that limits statistical utility. For example, the 'Weighing' question (Fairburn et al., 2008, p. 33), which measures the frequency of weighing behaviour, provides a unique opportunity to identify two opposing types of problematic weighing behaviour (i.e. weighing oneself too frequently or avoiding weighing oneself altogether). The reliable and valid measurement of both extremes on a continuum is particularly important to ED because often both extremes can be pathological (e.g. binge eating vs. restricting and body checking vs. body avoidance). In general, the EDE is an excellent tool for assessing opposing poles of ED psychopathology. However, the 'Weighing' item is an exception because it asks the assessor to indicate the frequency of weighing, but specifies that if the respondent has been avoiding weighing, then the item should be rated as '777'. Despite providing clinically useful data, a score of '777' cannot be used in conjunction with the dimensional frequency scores. Thus, both extremes of the continuum are not represented by scores on this item. In summary, some clinically relevant ED pathology is being assessed by the EDE, but rated in such a way that prevents them from being reflected in the total score.

The next generation of eating disorder assessment

When the EDE was first published in the seminal 1987 paper, it was an extraordinary innovation. But has it withstood the test of time? In our opinion, not quite. Twenty-five years after its initial publication, we face a critical fork in the road of ED assessment: we can either substantially streamline the EDE or cultivate a brave new post-EDE world.

Streamlining the EDE

To maintain its status as the gold-standard measure of ED psychopathology, the EDE would require updating. For example, the EDE will need to integrate constructs relevant to *DSM-5* (e.g. lack of recognition of the seriousness of low body weight, pica, rumination, and ARFID) to retain its relevance in the field. Unfortunately, the EDE has already expanded to the point of being cumbersome and inefficient. And as the ED literature expands, the EDE cannot expand indefinitely. Thus, perhaps the most major modification to the EDE needs to be the structure of the EDE rather than the content.

One option would be to substantially streamline the measure by focusing on items relevant to the subscales and diagnostic algorithms. Underutilised and/or unvalidated items (e.g. pattern of eating, body composition, and control items) could be summarily eliminated. Future versions of the EDE could require a higher burden of proof (e.g. using newer scale development paradigms such as item response theory) before adding clinically interesting but largely untested items that a generation of research assistants will be required to administer. (Just because we walked to school uphill through the snow both ways, does not mean they should, too.) The EDE could be further simplified by integrating the shape and weight items. For example, instead of assessing dissatisfaction with weight and dissatisfaction with shape separately, strategically using 'and/or' could

effectively reduce the complexity of the assessment. This modification is supported by factor analyses that have failed to distinguish between items relating to weight vs. shape.

A second option might be to replace the EDE with the EDE-Q, at least for the assessment of some constructs. For example, the EDE subscale scores correlate with those on the EDE-Q in the .70 range and frequency scores for self-induced vomiting and laxative misuse correlate in the .90 range (Berg et al., 2011). Thus, the EDE-Q could be used to assess the subscale items as well as the frequency of purging behaviours. Given that there is less consistency between scores on the EDE and EDE-Q for the assessment of binge eating (Berg et al., 2011) and concerns that lay definitions of binge eating are not consistent with *DSM* definitions (Beglin & Fairburn, 1992), assessors could elect to use the EDE to assess binge eating. The viability of replacing the EDE with the EDE-Q is further supported by research demonstrating reasonable diagnostic concordance (Berg, Stiles-Shields et al., 2012) and comparable latent structures (Berg, Swanson et al., 2013).

A third option may be to change the format of the EDE such that it becomes a module-based assessment. In this format, the EDE items would be grouped into categories and each category of items would form a module. The modules could then be used in any combination and in any order. For example, all the subscale items could form one module, all the items relevant to diagnostic algorithms could form a second module, and all the remaining items could form a third module. Alternatively, modules could be based on specific constructs (e.g. body dissatisfaction and restraint) or created for specific groups of users (e.g. researchers and clinicians). Assessors could then pick and choose which modules to conduct based on which constructs they are interested in measuring or how they want to use the assessment (e.g. dimensionally or diagnostically). This increase in flexibility would translate to an increase in efficiency – both in terms of time and cost – and lower assessor/respondent burden.

A brave, new post-EDE world

A more revolutionary option would be to retire the EDE as the gold-standard measure of ED psychopathology. Under this proposal, the EDE would no longer be considered as the necessary prerequisite for a well-conducted study. Instead, this seminal interview could be viewed as part of an ever-growing menu of ED assessment options. Indeed, several briefer but more comprehensive measures are currently under development.

In the realm of diagnosis, one new assessment instrument is the Eating Disorder Assessment-5 (EDA-5) (Glasofer et al., 2012). It is a semi-structured interview designed to assess the presence or absence of *DSM-5* feeding and eating disorders. Early versions of the measure, administered at three ED treatment sites, have demonstrated good test-retest reliability and adequate concordance with EDE diagnoses. Preliminary data suggest that it is acceptable to patients and takes significantly less time to administer (i.e. just 15–20 minutes) than the EDE. Development of an electronic application of EDA-5 is currently underway.

When it comes to continuous measures of eating pathology, Forbush and colleagues (2013) have recently developed a self-report measure called as the Eating Pathology Symptoms Inventory. Importantly, this comprehensive multi-dimensional measure was constructed empirically using an initial validation sample of more than 1500 subjects from clinical, university, and community samples. Confirmatory factor analysis revealed an eight-factor structure that overcomes many of the construct validity limitations inherent to the EDE. For example, Restricting and Cognitive Restraint are measured separately; Body Dissatisfaction combines shape- and weight-related items; Muscle Building highlights components of body image disturbance particularly relevant to males; and the remaining scales (i.e. Binge Eating, Excessive Exercise, Purging, and Negative Attitudes towards Obesity) provide comprehensive coverage of the remaining aspects

of eating pathology. Scale scores are invariant across sex and weight categories and show excellent convergent and discriminant validity.

Conclusion

Twenty-five years after its initial publication, the EDE – the field's gold-standard assessment of ED psychopathology – has many critical limitations. Just as modern economies have diversified based on evolving financial needs, so too must we as a field consider how to optimise the assessment of an increasingly diverse array of patients, syndromes, and symptoms. The EDE could be substantially streamlined or even largely retired. And before you gasp in horror, please remember this: although gold served as the original basis for many modern monetary systems, nowhere in the world is the gold standard currently in use.

Acknowledgements

The authors would like to thank the Hilda and Preston Davis Foundation, as well as the Robert Wood Johnson Foundation Health & Society Scholars programme for their financial support.

Note

1. e.g. '0' = 0 days, '1' = 1–5 days, '2' = 6–12 days, '3' = 13–15 days, '4' = 17–22 days, '5' = 23–27 days, and '6' = 28 days.

References

Aardoom, J. J., Dingemans, A. E., Landt, M. C. T. S. O., & Van Furth, E. F. (2012). Norms and discriminative validity of the eating disorder examination questionnaire (EDE-Q). *Eating Behaviors*, *13*, 305–309.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders, Fifth edition (DSM-5)*. Washington, DC: Author.

Becker, A. E., Thomas, J. J., Bainivualiku, A., Richards, L., Navara, K., Roberts, A. L., … Striegel-Moore, R. H. (2010). Validity and reliability of a Fijian translation and adaptation of the eating disorder examination questionnaire. *International Journal of Eating Disorders*, *43*, 171–178.

Beglin, S. J., & Fairburn, C. G. (1992). What is meant by the term "binge"? *American Journal of Psychiatry*, *149*, 123–124.

Berg, K. C. (2010). *A study of the validity of the Eating Disorder Examination* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis, MN.

Berg, K. C., Crosby, R. D., Cao, L., Peterson, C. B., Engel, S. G., Mitchell, J. E., & Wonderlich, S. A. (2013). Facets of negative affect prior to and following binge-only, purge-only, and binge/purge events in women with bulimia nervosa. *Journal of Abnormal Psychology*, *122*, 111–118.

Berg, K. C., Peterson, C. B., & Crow, S. (2009, September). *Assessment of binge eating: The concordance between the EDE and daily food records*. Poster session presented at the 15th Eating Disorders Research Society Annual Meeting, Brooklyn, NY.

Berg, K. C., Peterson, C. B., Frazier, P., & Crow, S. J. (2011). Convergence of scores on the interview and questionnaire versions of the eating disorder examination: A meta-analytic review. *Psychological Assessment*, *23*, 714–724.

Berg, K. C., Peterson, C. B., Frazier, P., & Crow, S. J. (2012). Psychometric evaluation of the Eating Disorder Examination and Eating Disorder Examination-Questionnaire: A systematic review of the literature. *International Journal of Eating Disorders*, *45*, 428–438.

Berg, K. C., Stiles-Shields, E. C., Swanson, S. A., Peterson, C. B., Lebow, J., & Le Grange, D. (2012). Diagnostic concordance of the interview and questionnaire versions of the Eating Disorder Examination. *International Journal of Eating Disorders*, *45*, 850–855.

Berg, K. C., Swanson, S. A., Stiles-Shields, E. C., Eddy, K. T., Peterson, C. B., & Le Grange, D. (2013). Response patterns on interview and questionnaire versions of the Eating Disorder Examination and their impact on latent structure analyses. *Comprehensive Psychiatry, 54*, 506–516.

Byrne, S. M., Allen, K. L., Lampard, A. M., Dove, E. R., & Fursland, A. (2010). The factor structure of the eating disorder examination in clinical and community samples. *International Journal of Eating Disorders*, *43*, 260–265.

Cooper, Z., Cooper, P. J., & Fairburn, C. G. (1989). The validity of the eating disorder examination and its subscales. *The British Journal of Psychiatry*, *154*, 807–812.

Cooper, Z., & Fairburn, C. (1987). The eating disorder examination: A semi-structured interview for the assessment of the specific psychopathology of eating disorders. *International Journal of Eating Disorders*, *6*, 1–8.

Crow, S. J., Agras, W., Halmi, K., Mitchell, J. E., & Kraemer, H. C. (2002). Full syndromal versus subthreshold anorexia nervosa, bulimia nervosa, and binge eating disorder: A multicenter study. *The International Journal of Eating Disorders*, *32*, 309–318.

Dalle Grave, R., Calugi, S., Doll, H. A., & Fairburn, C. G. (2013). Enhanced cognitive behaviour therapy for adolescents with anorexia nervosa: An alternative to family therapy. *Behaviour Research and Therapy*, *51*, R9–12.

Darcy, A. M., Doyle, A. C., Lock, J., Peebles, R., Doyle, P., & Le Grange, D. (2012). The eating disorders examination in adolescent males with anorexia nervosa: How does it compare to adolescent females? *International Journal of Eating Disorders*, *45*, 110–114.

Eddy, K. T., Le Grange, D., Crosby, R. D., Hoste, R. R., Doyle, A. C., Smyth, A., & Herzog, D. B. (2010). Diagnostic classification of eating disorders in children and adolescents: How does DSM-IV-TR compare to empirically-derived categories? *Journal of the American Academy of Child & Adolescent Psychiatry, 49*, 277–287.

Engel, S. G., Wonderlich, S. A., Crosby, R. D., Mitchell, J. E., Crow, S., Peterson, C. B., … Gordon, K. H. (2013). The role of affect in the maintenance of anorexia nervosa: Evidence from a naturalistic assessment of momentary behaviors and emotion. *Journal of Abnormal Psychology*, *122*, 709–719.

Fairburn, C. G. (2008). *Cognitive behavior therapy and eating disorders*. New York: Guilford Press.

Fairburn, C. G., & Cooper, Z. (1993). The eating disorder examination. In C. G. Fairburn, & G. T. Wilson (Eds.), *Binge eating: Nature, assessment, and treatment* (12th ed., pp. 317–360). New York, NY: Guilford.

Fairburn, C. G., Cooper, Z., Doll, H. A., O'Connor, M. E., Bohn, K., Hawker, D. M., … Palmer, R. L. (2009). Transdiagnostic cognitive-behavioral therapy for patients with eating disorders: A two-site trial with 60-week follow-up. *American Journal of Psychiatry*, *166*, 311–319.

Fairburn, C. G., Cooper, Z., Doll, H. A., O'Connor, M. E., Palmer, R. L., & Dalle Grave, R. (2013). Enhanced cognitive behaviour therapy for adults with anorexia nervosa: A UK-Italy study. *Behaviour Research and Therapy*, *51*, R2–8.

Fairburn, C. G., Cooper, Z., & O'Connor, M. (2008). Eating disorder examination, edition 16.0D. In C. G. Fairburn (Ed.), *Cognitive behavior therapy and eating disorders* (pp. 265–308). New York, NY: Guilford Press.

Farchaus Stein, K., & Corte, C. M. (2003). Ecological momentary assessment of eating-disordered behaviors. *International Journal of Eating Disorders*, *34*, 349–360.

Fichter, M. M., Elton, M., Engel, K., Meyer, A. E., Mall, H., & Poustka, F. (1991). Structured Interview for Anorexia and Bulimia Nervosa (SIAB): Development of a new instrument for the assessment of eating disorders. *International Journal of Eating Disorders*, *10*, 571–592.

Forbush, K. T., Wildes, J. E., Pollack, L. O., Dunbar, D., Luo, J., Patterson, K., … Watson, D. (2013). Development and validation of the eating pathology symptoms inventory (EPSI). *Psychological Assessment*, *25*, 859–878.

Friborg, O., Reas, D. L., Rosenvinge, J. H., & Ro, O. (in press). Core pathology of eating disorders as measured by the eating disorder examination questionnaire (EDE-Q): The predictive role of a nested general (g) and primary factors. *International Journal of Methods in Psychiatric Research*.

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., & 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.

Glasofer, D. R., Sysko, R., Attia, E., Crosby, R. D., Hildebrandt, T., Mitchell, J. E., … Walsh, B. T. (2012, September). *Eating disorder assessment for DSM-5: A structured clinical interview for DSM-5 eating disorders*. Paper presented at the meeting of the Eating Disorder Research Society, Porto, Portugal.

Haedt-Matt, A. A., & Keel, P. K. (2011). Revisiting the affect regulation model of binge eating: A meta-analysis of studies using ecological momentary assessment. *Psychological Bulletin*, *137*, 660–681.

Loeb, K. L., Pike, K. M., Walsh, B. T., & Wilson, G. T. (1994). Assessment of diagnostic features of bulimia nervosa: Interview versus self-report format. *International Journal of Eating Disorders*, *16*, 75–81.

Lowinger, K., Griffiths, R. A., Beumont, P. J., Scicluna, H., & Touyz, S. W. (1999). Fluid restriction in anorexia nervosa: A neglected symptom or a new phenomenon? *International Journal of Eating Disorders*, *26*, 392–396.

Ludwig, D. S., Peterson, K. E., & Gortmaker, S. (2001). Relation between consumption of sugar-sweetened drinks and childhood obesity: A prospective, observational analysis. *The Lancet*, *357*, 505–508.

Masheb, R. M., Roberto, C. R., & White, M. A. (in press). Nibbling and picking in obese patients with binge eating disorder. *Eating Behaviors*.

Mitchell, J. E., Agras, W. S., Crow, S. J., Halmi, K., Fairburn, C. G., Bryson, S., & Kraemer, H. (2011). Stepped care and cognitive-behavioural therapy for bulimia nervosa: A randomised trial. *The British Journal of Psychiatry*, *128*, 391–398.

O'Connor, M. (2009, September 14). *Oxford EDE 16.0 training*o.Warneford Hospital, Oxford University, United Kingdom.

Otto, J. (2011). *Never drink calories: Theories of an anorexic*. Minneapolis, MN: Mill City Press.

Reas, D. L., Wisting, L., Kapstad, H., & Lask, B. (2012). Nibbling: Frequency and relationship to BMI, pattern of eating, and shape, weight and eating concerns among university women. *Eating Behaviors*, *13*, 65–66.

Roberto, C. A., Grilo, C. M., Masheb, R. M., & White, M. A. (2010). Binge eating, purging, or both: Eating disorder psychopathology findings from an Internet community survey. *International Journal of Eating Disorders*, *43*, 724–731.

Roberto, C. A., Steinglass, J., Mayer, L., Attia, E., & Walsh, B. T. (2008). The clinical significance of amenorrhea as a diagnostic criterion for anorexia nervosa. *International Journal of Eating Disorders*, *41*, 559–563.

Rosen, J. C., Vara, L., Wendt, S., & Leitenberg, H. (1990). Validity studies of the eating disorder examination. *International Journal of Eating Disorders*, *9*, 519–528.

Smyth, J. M., Wonderlich, S. A., Heron, K. E., Sliwinski, M. J., Crosby, R. D., Mitchell, J. E., & Engel, S. G. (2007). Daily and momentary mood and stress are associated with binge eating and vomiting in bulimia nervosa patients in the natural environment. *Journal of Consulting and Clinical Psychology*, *75*, 629–638.

Sobell, L. C., Maisto, S. A., Sobell, M. B., & Cooper, A. M. (1979). Reliability of alcohol abusers' self-reports of drinking behavior. *Behavior Research & Therapy*, *17*, 157–160.

Sobell, M. B., Sobell, L. C., Klajner, F., Pavan, D., & Basian, E. (1986). The reliability of a timeline method for assessing normal drinker college students' recent drinking history: Utility for alcohol research. *Addictive Behaviors*, *11*, 149–161.

Stice, E., Fisher, M., & Lowe, M. R. (2004). Are dietary restraint scales valid measures of acute dietary restriction? Unobtrusive observational data suggest not. *Psychological Assessment*, *16*, 51–59.

Stice, E., Telch, C. F., & Rizvi, S. L. (2000). Development and validation of the eating disorder diagnostic scale: A brief self-report measure of anorexia, bulimia, and binge-eating disorder. *Psychological Assessment*, *12*, 123–131.

Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*, *16*, 199–202.

Striegel-Moore, R. H., Franko, D. L., Thompson, D., Barton, B., Schreiber, G. B., & Daniels, S. R. (2005). An empirical study of the typology of bulimia nervosa and its spectrum variants. *Psychological Medicine*, *35*, 1563–1572.

Swanson, S. A., Crow, S. J., Le Grange, D., Swendson, J., & Merikangas, K. R. (2011). Prevalence and correlates of eating disorders in adolescents: Results from the national comorbidity survey replication adolescent supplement. *Archives of General Psychiatry*, *68*, 714–723.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics* (4th ed.). Boston, MA: Allyn and Bacon.

Teasdale, J. D., & Fogarty, S. J. (1979). Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology*, *88*, 248–257.

Thomas, J. J., Roberto, C. A., & Brownell, K. D. (2009). Eighty-five per cent of what? Discrepancies in the weight cut-off for anorexia nervosa substantially affect the prevalence of underweight. *Psychological Medicine*, *39*, 833–843.

Thomas, J. J., Vartanian, L. R., & Brownell, K. D. (2009). The relationship between eating disorder not otherwise specified (EDNOS) and officially recognized eating disorders: Meta-analysis and implications for DSM. *Psychological Bulletin*, *135*, 407–433.

Wildes, J. E., & Marcus, M. D. (2011). Development of emotion acceptance behavior therapy for anorexia nervosa: A case series. *International Journal of Eating Disorders*, *44*, 421–427.

Wonderlich, S. A., Peterson, C. B., Smith, T. L., Klein, M., Mitchell, J. E., Crow, S. J., & Engel, S. G. (2009). Integrative cognitive–affective therapy for bulimia nervosa. In Carlos M. Grilo & James E. Mitchell (Eds.), *The treatment of eating disorders: A clinical handbook* (pp. 317–338). New York: Guilford.