

New Horizons in Measurement: a Review of Novel and Innovative Approaches to Eating-Disorder Assessment

Kelsie T. Forbush¹ · Sara R. Gould² · Danielle A. N. Chapa¹ · Brittany K. Bohrer¹ ·
Kelsey E. Hagan¹ · Kelsey E. Clark¹ · Daria A. Sorokina¹ · Victoria L. Perko¹

Published online: 11 September 2017
© Springer Science+Business Media, LLC 2017

Abstract

Purpose of Review Eating disorders are serious mental-health concerns that will affect over 30 million individuals in the USA at some point in their lives. Eating disorders occur across the lifespan, in a variety of ethnicities and races, in both men and women, and across the socioeconomic spectrum. Given the prevalence and severity of eating disorders, it is important that clinicians and researchers have access to appropriate assessment tools to aid in the early identification and treatment referral, differential diagnosis, treatment planning, and progress monitoring, and to ensure valid research findings. In this review, we describe novel and innovative assessment tools that were developed within the past 5 years for utilization in research and/or clinical practice with individuals with eating disorders.

Recent Findings We identified six multidimensional assessments for eating disorders, all of which can be administered online (with some also offering paper-and-pencil versions). Strengths of the measures included good internal consistency, test-retest reliability, and convergent validity. However, in part, due to problematic scale construction methods, certain scales had poor discriminant validity and most were developed and validated in mostly female samples.

Summary There are promising new eating disorder measures from which to choose; however, many measures continue to be limited by poor discriminant validity and need additional

validation prior to incorporation into routine research and clinical practice.

Keywords Eating disorders · Disordered eating · Assessment · Measurement · Reliability · Validity

Introduction

Eating disorders are more common than once believed, with prevalence estimates from community- and population-based studies indicating that approximately 2.5% of boys and 13–15% of girls will suffer from an eating disorder by age 17 [1, 2]. Eating disorders are serious mental-health conditions that are associated with a high mortality rate [3] and substantial medical and psychiatric morbidity [4, 5]. Age-of-onset typically occurs in childhood or adolescence [6], and for many people, eating disorders become chronic, long-lasting conditions [7, 8]. Eating disorders have been documented across the life span [8, 9], among individuals of a variety of ethnic and national backgrounds [10–12], in both men and women [13, 14], and across the socioeconomic spectrum [15]. Therefore, it is important that clinicians and researchers have access to eating-disorder assessment tools that have been validated in diverse populations to facilitate early identification (for prevention and treatment referral purposes), accurate diagnosis, and effective treatment planning, and to ensure valid research findings.

From a treatment perspective, the importance of proper assessment cannot be overstated. Research has shown that one of the only significant moderators of treatment outcomes for clients with an eating disorder is rapid, early change that occurs between the second and eighth treatment week [16–18]. In order to determine if patients are on a good treatment trajectory, repeated assessments of outcomes are

This article is part of the Topical Collection on *Eating Disorders*

✉ Kelsie T. Forbush
kforbush@ku.edu

¹ Department of Psychology, University of Kansas, Fraser Hall, 1415 Jayhawk Boulevard, Lawrence, KS 66045, USA

² Children's Mercy Kansas City, Kansas City, MO, USA

necessary. Moreover, a large literature outside of the eating-disorder field has shown that clinicians who are provided with objective feedback from assessments of their clients' progress are able to treat their clients with fewer therapy sessions, without reducing positive therapeutic outcomes, and that their clients show less deterioration of treatment gains after termination [19–22]. Thus, not only are eating-disorder measures important for accurately describing patients' symptoms across time, but assessment may also provide additional treatment benefits above-and-beyond therapeutic intervention, consistent with the principles of Therapeutic Assessment [23].

Overview of Current Review

The aim of the current paper was to review novel and innovative assessment tools that were developed within the past 5 years for utilization in research and/or clinical practice with individuals with eating disorders. Although several well-validated eating-disorder assessments have been developed to measure specific eating disorders (e.g., binge-eating disorder [24]) and core cognitive aspects of eating-disorder psychopathology (e.g., body image acceptance [25]) or to measure behaviors for a specific subgroup of individuals, such as athletes [26], we limited our review to “all-in-one” measures that were developed to assess multiple domains and symptoms that cut across eating-disorder diagnoses and populations.

After describing our method for identifying assessment tools, we discuss each measure in terms of its purpose and content, appropriate uses and populations, scale development process, psychometric properties (including evidence for reliability and validity), and strengths and limitations. We conclude our review with a comparison of the advantages and disadvantages of the measures we reviewed and provide suggestions for future research in the area of eating-disorder assessment.

Method

This systematic literature review was conducted in accordance with the Preferred Reporting Items for Systematic Review and Meta-Analysis statement (PRISMA) [27]. Throughout the current review, our definitions of reliability and validity were based on standard usage within the assessment field [28]. Because standard usage of psychometric terms within the eating-disorder field were, in some cases, inconsistent with definitions from the broader assessment literature, we refer interested readers to Table 1 for a list of terms and definitions that were used in the current review.

Table 1 Psychometric terms and definitions

Psychometric term	Definition
Reliability	The ability of an instrument to perform in a consistent manner in which scores are reflective of some true state of the variable being assessed.
Internal Consistency	The extent to which all items of an instrument or scale are correlated.
Test-retest reliability (dependability)	The ability of an instrument to perform in a constant manner when the variable of interest has remained stable over periods of time during which true change is unlikely to have occurred (usually over periods of 1 day to 4 weeks).
Stability	The ability of an instrument to demonstrate constancy over time; a quantification of both measurement error and true change. Typically assessed over periods of time in which true change is likely (e.g., greater than 1 month).
Validity	The extent to which an instrument measures the theoretical construct that it intends to measure.
Construct validity	The extent to which an operationalization of a variable relates to other measures in a manner that is consistent with the theoretical concept it intends to capture.
Convergent	The extent to which two measures of theoretically related constructs correspond. Higher values indicate good convergent validity between measures.
Discriminant	The extent to which two measures of theoretically unrelated constructs do not correspond. Lower values indicate good discriminant validity between measures.
Criterion-related validity	The extent to which an instrument corresponds with a predictable outcome of interest; divided into concurrent validity and predictive validity (see below).
Concurrent	The ability of an instrument to accurately predict an outcome of interest based on validation that occurs at the same time point.
Predictive	The ability of an instrument to accurately predict an outcome of interest based on validation that occurs at a later time point.
Content validity	The extent to which an instrument reflects all aspects of the content domain it attempts to measure.
Sensitivity	The ability of an instrument to correctly designate an individual with a characteristic of interest as a member of the appropriate (correct) category.
Specificity	The ability of an instrument to correctly designate an individual without a characteristic of interest as a member of the appropriate category.

Please refer to DeVellis [28] for additional information on reliability and validity definitions

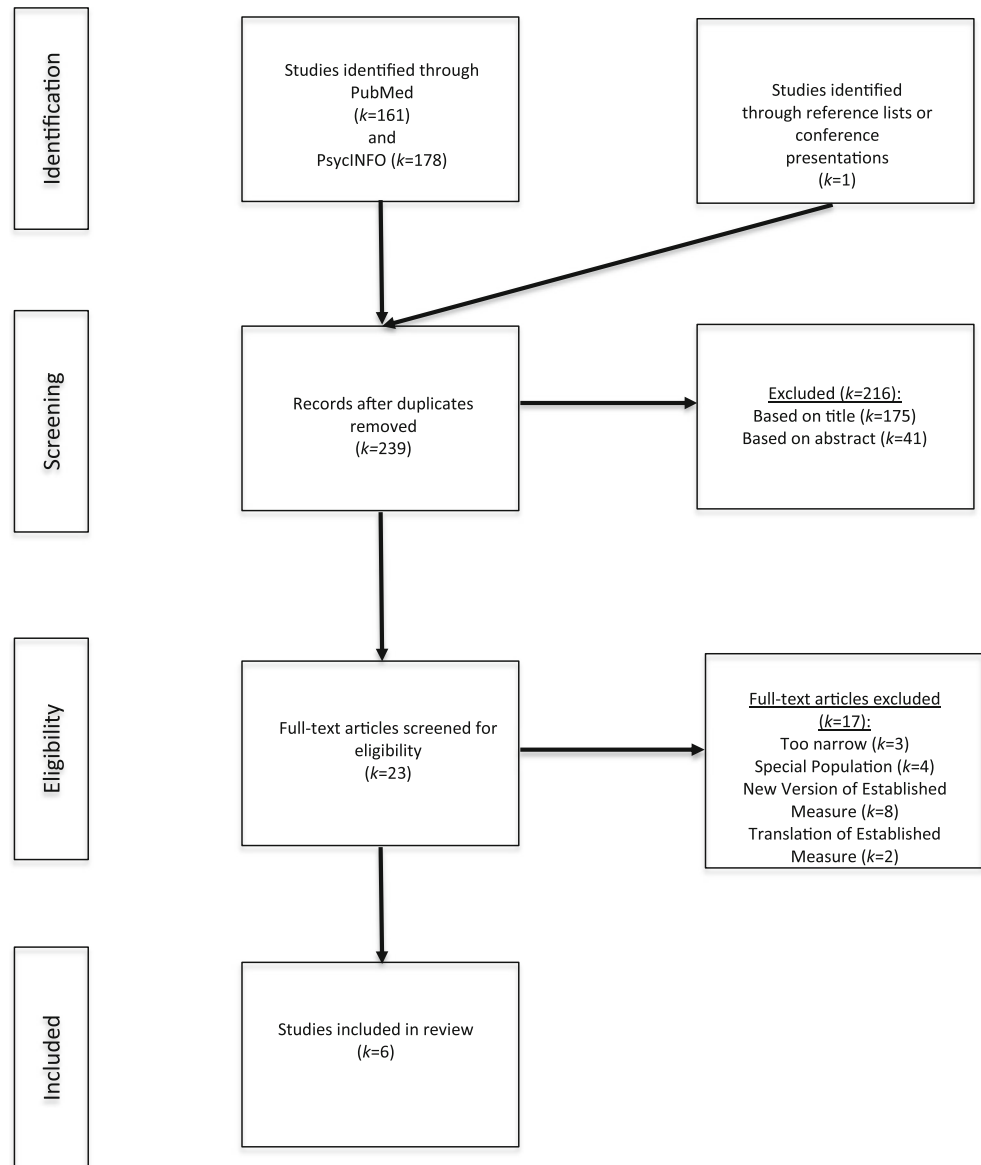
Literature Search

Several methods were used to identify relevant articles. First, an electronic database search was conducted from articles

published from January 1, 2012, through March 14, 2017 using PsycINFO and PubMed. The following combination of terms was utilized: (“eating disorder*” OR anore* OR bulim* OR “binge eating disorder*” OR “night eating syndrome” OR “avoidant restrictive food intake disorder” OR “purging disorder” OR “orthorexia” OR OSFED OR EDNOS OR ARFID OR “disordered eating”) AND (assessment OR “self-report” OR interview OR questionnaire OR measure OR inventory). Searches were limited to titles of manuscripts. In addition to electronic database searches, reference lists of manuscripts retrieved from electronic databases were manually reviewed for relevant studies not identified through electronic database search. Finally, the authors of assessment tools identified in our search were contacted to determine if additional pertinent studies had been missed or were “in press.”

Two authors independently completed an electronic database search (B.B. and D.C.) and the second author (S.G.) resolved discrepancies regarding the inclusion of articles. Our database search included several steps: First, information from articles identified from PsycINFO and PubMed were downloaded into an Excel spreadsheet and duplicate articles were removed. Second, manuscript titles were reviewed and titles that did not pertain to the assessment of disordered eating were removed. Third, articles pertaining to the assessment of a specific eating-disorder symptom, designed for use in a specific population, or were modifications or translations of existing measures were removed. Finally, remaining abstracts were screened and full-text articles of studies were retrieved and reviewed. A PRISMA flowchart of the manuscript review process is presented in Fig. 1.

Fig. 1 PRISMA flowchart of study selection. *Note:* The figure reflected the original publications for each included measure. We also identified two additional measures further validating the EPSI. The authors of the other measures responded to our queries that there are no additional publications for the other measures at this time



Eligibility Criteria

Studies were included in the current review if they (1) provided information on a novel assessment of eating disorders, (2) were written in English or English translation available, (3) assessed eating-disorder symptoms comprehensively, and (4) were published in peer-reviewed journals (e.g., publications from book chapters, theses, and dissertations were excluded). Updated versions of previously published assessments (such as the Eating Disorder Inventory-3; [29]) were not included in this review nor were assessments that were designed to assess a limited symptom set (e.g., binge eating) or population (e.g., males or athletes).

Results

A total of 238 unique studies were identified through our data search procedure. Of these, five studies met inclusion criteria and were included in our review (see Fig. 1). Additionally, we identified one “in press” article by contacting an author of a measure that had been presented at academic conferences during 2012–2017. Only the initial publication of this measure (the Eating Pathology Symptom Inventory) is reflected in Fig. 1 (see Fig. 1 note). Below, we describe each of these six novel assessments.

The Clinical and Research Inventory for Eating Disorders

Purpose and Content The Clinical and Research Inventory for Eating Disorders (CR-EAT; [30••]) is a 63-item self-report measure designed to provide clinicians and researchers with a comprehensive overview of the cognitive and behavioral features relevant to the development, maintenance, prevention, and treatment of eating disorders. The CR-EAT has been translated into multiple languages (English, German, French, Portuguese, Spanish, Hungarian, Romanian, and Czech) and is freely available for research purposes by contacting the authors (see [30••]).

The CR-EAT has 11 subscales grouped into three global factors: *eating behavior disturbance*, *affective/cognitive impairment*, and *perfectionism*. The 11 subscales include *weight preoccupation* (obsession and worry about weight and shape, unrealistic concerns about weight gain), *body embarrassment* (shame and dissatisfaction surrounding one’s body), *restrained eating behavior* (maintenance of strict rules to control caloric intake), *societal expectations of weight and shape* (belief that social acceptance hinges on one’s shape and weight), *harmful weight regulation* (willingness to use unhealthy weight loss methods), *mood dysregulation* (depressive feelings and deficits in emotional self-awareness), *affect-regulatory eating* (eating as a means of emotion regulation), *self-esteem* (confidence in one’s self-worth), *concerns about*

negative evaluation (concern about others’ opinions about oneself), *perfectionism: personal expectations* (holding oneself to perfectionistic expectations), and *perfectionism: familial expectations* (perceived familial/parental perfectionistic standards).

Appropriate Uses The CR-EAT is intended to be an efficient, multidimensional eating disorder assessment for use in mental health settings. The CR-EAT has not been tested in individuals younger than age 16 or in clinical samples of men. According to Moessner et al. [30••], women score higher than men on all three global scales, and on eight of the 11 subscales (weight preoccupation, body embarrassment, restrained eating behavior, harmful weight regulation, mood dysregulation, affect-regulatory eating, self-esteem, and concerns about negative social evaluation).

Scale Development The CR-EAT was developed using a factor-analytic approach. Following a comprehensive literature review and expert consultation, a rational strategy was used to develop the original item pool. The initial item pool was administered once online to a large nonclinical sample ($N = 1406$; 74.3% female) and twice via paper-and-pencil questionnaires to another nonclinical sample ($N = 220$; 71.4% female). The initial 100 items were also administered online to a subset of the clinical sample with eating disorders ($N = 61$; 100% female). Following principle components analysis with varimax rotation, the initial item pool was reduced to a final set of 63 items that comprised 11 lower-order scales and three global scales.

To evaluate the psychometric properties of the CR-EAT, additional paper-and-pencil questionnaires including the Eating Disorder Inventory-2 (EDI-2; [31]), the Short Evaluation of Eating Disorders (SEED; [32]), and the Weight Concerns Scale (WCS; [33]) were administered repeatedly to a nonclinical sample ($N = 220$), who completed measures twice over a 4-week period.

Reliability Internal consistency (as assessed by Cronbach’s α) ranged from 0.62 for harmful weight regulation to 0.95 for the CR-EAT total score. Test-retest reliability over a 4-week period (intraclass correlation coefficients, ICC) ranged from 0.83 for self-esteem to 0.97 for CR-EAT total score.

Validity Convergent validity was examined via correlations among the CR-EAT with the EDI-2, SEED, and WCS. Some CR-EAT subscale scores correlated strongly with scores on measures of theoretically similar constructs (e.g., CR-EAT body embarrassment and EDI-2 body dissatisfaction). However, certain CR-EAT subscales showed poor discriminant validity as evidenced by high correlations with theoretically divergent constructs (e.g., $r = 0.74$ for the CR-EAT binge eating and EDI-2 body dissatisfaction), while simultaneously

showing low convergent correlations with related constructs (e.g., $r = 0.36$ for the CR-EAT binge eating and SEED Bulimia Nervosa Total Severity Index).

Data from the clinical sample of participants with eating disorders were used to test criterion-related validity via sensitivity-specificity analyses that controlled for sex. Results showed that CR-EAT scales distinguished between clinical and nonclinical sample participant groups (note: the clinical population in the study was exclusively female).

Strengths and Limitations Moessner and colleagues [30••] developed and validated the CR-EAT questionnaire with the field of e-mental health in mind as a particular beneficiary. Interest in Internet-based eating-disorder intervention and prevention programs is growing [34], given the many advantages of online approaches (e.g., anonymity, low cost). While online treatment programs typically rely on self-report measures to assess change over time, they often use measures that have not been validated for online administration. The literature regarding the generalizability and transferability of administration medium from paper-and-pencil to online is somewhat limited for eating-disorder assessments. Studies have found excellent comparative validity between online and paper versions of some assessments, including the Night Eating Questionnaire [35] and the Weight Concerns Scale [36]. A systematic review of studies comparing online and paper-and-pencil psychiatric assessments [37] found that digital adaptations of paper self-report measures generally demonstrated high reliability. However, Alfonsso et al. [37] noted some exceptions and, therefore, recommend that reliability be tested for individual measures and not assumed to be present in every case. Many popular existing eating-disorder measures have not been validated for online assessment, so the CR-EAT represents a novel development.

To our knowledge, no other studies, to date, have examined the psychometric properties of the CR-EAT, which was confirmed by Moessner (personal communication, April 2017). While the initial analyses of the CR-EAT showed evidence for strong reliability, there were substantial issues related to the convergent and discriminant validity of this measure. The current version of the CR-EAT may not afford optimal assessment of the intended constructs for clinical and research purposes. Additional analysis of the psychometric properties of the measure is warranted in order to provide data that could be used to improve the construct validity of the CR-EAT in future editions/revisions.

ED-15

Purpose and Content The ED-15 [38••] is a self-report measure that was designed for use in treatment settings to provide a session-by-session (i.e., weekly) outcome assessment of eating-disorder attitudes and behaviors. The ED-15 has 10

attitudinal items which comprise two factors: (1) *weight and shape concerns* (concerns with body shape and weight) and (2) *eating concerns* (concerns with the types or amounts of food consumed). The ED-15 includes five additional items that assess the frequency of binge-eating and inappropriate compensatory behaviors (e.g., self-induced vomiting, laxative use, restricting, and excessive exercise) over the past week. Although the content and factor structure of the ED-15 appears extremely similar to the Eating Disorder Examination-Questionnaire (EDE-Q; [39]), the ED-15 is a unique measure that was developed independently from the EDE-Q. A copy of the ED-15 can be found in the appendix of the development article [38••].

Appropriate Uses The ED-15 was designed to complement other eating-disorder assessments; thus, the ED-15 is not intended to be the only assessment of eating-disorder psychopathology used in a clinical setting. Rather, the ED-15 was developed to assist clinicians and researchers with tracking eating-disorder symptom change from week to week. Clinicians can use the ED-15 to assess whether their treatment plan effectively facilitated client change. ED-15 scores that increase or remain elevated may suggest that the clinician's treatment plan needs to be reevaluated or another intervention needs to be introduced. The ED-15 has been tested in adult men and women without eating disorders and in adult women with eating disorders. Future research is needed to validate the ED-15 in adolescents, children, and men with diagnosable eating disorders.

Scale Development The authors of the ED-15 wrote 16 items based on theory and their previous clinical experience in the treatment of eating disorders. The original, 16-item pool was administered to a nonclinical sample of university men and women and staff members between the ages of 18 and 71 years ($N = 531$; 82.5% female). The factor structure of the initial item pool was assessed using principle component analysis in women ($n = 438$). One attitudinal item was excluded from the final item pool because it cross-loaded onto both factors. (It is important to note that the factor structure of the ED-15 was not tested in men). During scale development, the ED-15 was administered to women with a self-reported eating disorder ($N = 63$; mean age = 28.7) and women who were diagnosed with bulimia nervosa based on the EDE interview ($N = 33$; mean age = 30.8).

Reliability and Sensitivity to Change Among a sample of university students and staff members without an eating disorder, internal consistency reliability for the ED-15 was good to excellent (Cronbach's alpha ranged from 0.80 for eating concerns to 0.94 for weight and shape concerns). Test-retest reliability over a mean of 18 days in a subset of the nonclinical sample of men and women ($n = 149$; 86.6% female) ranged

from 0.85 to 0.92. Among women who were diagnosed with bulimia nervosa who completed the ED-15 twice over a mean of 7 days ($n = 23$), test-retest reliability ranged from 0.79 for weight and shape concerns to 0.81 for eating concerns. In addition to the dependability of the ED-15, it was also sensitive to change, given that women who were receiving treatment for bulimia nervosa had significant decreases in their ED-15 scores over a 10-week period.

Validity Mean scores on the ED-15 were significantly higher for individuals with a self-reported eating disorder compared to individuals without a self-reported eating disorder, demonstrating evidence for criterion-related validity. The ED-15 attitudinal and behavioral items showed evidence for strong convergent validity with the EDE-Q in a nonclinical sample. For example, the ED-15 weight and shape concern subscale was significantly correlated with all of the EDE-Q subscales (r 's ranged from 0.55 to 0.88) but showed the highest correlations with the EDE-Q weight concern ($r = 0.86$) and EDE-Q shape concern ($r = 0.88$) subscales. Despite the strong convergent validity of the ED-15 with the EDE-Q, discriminant validity was problematic. For example, the ED-15 weight and shape subscale was substantially correlated with a measure of generalized anxiety ($r = 0.52$) and more highly correlated with a measure of depression ($r = 0.63$) than with the EDE-Q restraint subscale ($r = 0.55$).

Strengths and Limitations Although previous measures have been developed to assess week-to-week eating-disorder symptom change in treatment settings (e.g., the Change in Eating Disorder Symptoms Scale [40]), the ED-15 is notable as one of only a few eating-disorder self-report measures to assess symptom change on a week-to-week basis. The ED-15 is, therefore, an important new tool for tracking patient outcomes in clinical research trials and for evaluating clients' progress throughout treatment. The ED-15 has demonstrated strong internal consistency and test-retest reliability, suggesting that observed changes in scale scores are likely attributable to true change rather than measurement error (i.e., instability).

An important limitation is that the ED-15 has poor discriminant validity from generalized anxiety and depression. Poor discriminant validity could be due to the wording of certain ED-15 items, which may have inadvertently tapped into negative affectivity (e.g., "...worried about losing control ..." and "...depressed about my body shape"). As noted by Clark and Watson [41], including words that reflect negative affectivity should be avoided when creating scales, unless the goal of the measure is to assess depression or anxiety. At present, it is unclear whether changes in ED-15 scales reflect changes in eating-disorder behaviors and cognitions or depression (or both).

Eating Disorders Assessment for DSM-5

Purpose and Content The Eating Disorder Assessment for DSM-5 (EDA-5; [42••]) is a semistructured interview that assesses *Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition (DSM-5; [43])* feeding and eating-disorder criteria. The EDA-5 was developed to create a diagnostic instrument that could be used by researchers and clinicians with limited training. The EDA-5 can be administered via a paper-and-pencil administration or an electronic application ("app"). The interviewer asks questions according to a built-in skip logic algorithm. Because skip-out rules can be difficult to follow and can lead to coding errors when administered via the traditional paper-and-pencil format, the authors created an electronic application of the EDA-5 that automates the process of following skip-out rules. The electronic version of the EDA-5 is available at no cost at www.eda5.org.

Appropriate Uses The EDA-5 and the EDA-5 app were developed and validated in a treatment-seeking sample of adolescents and adults (ages ranged from 14 to 65). The authors suggested that the EDA-5 should be used in combination with other clinical information, when available (e.g., family reports and objective height and weight measurements), to assist in the provision of an accurate diagnosis.

Scale Development The authors of the EDA-5 used a DSM-5 symptom checklist to develop items that corresponded to each DSM-5 feeding and eating disorder. Items were created rationally, meaning that the authors wrote items to correspond with DSM-5 symptom criteria using their best judgment and experience in the field. The initial paper-and-pencil version of the EDA-5 was administered to individuals seeking or already receiving treatment at three tertiary care centers for eating disorders. Participants ($N = 64$; 89.1% women; 78.1% Caucasian) completed both the EDA-5 and Eating Disorder Examination (EDE; [44]) on the same day to assess concurrent validity. A subset of participants ($n = 21$) were randomized to complete the EDA-5 a second time approximately 7 to 14 days after the first interview to assess test-retest reliability. The EDA-5 app was also administered to a sample of individuals seeking or receiving treatment for an eating disorder ($N = 71$) along with other measures of eating-disorder psychopathology to test for concurrent validity. Concurrent validity can be expressed as a percentage of diagnostic agreement between two measures or with coefficient kappa, which expresses the extent that two measures arrive at the same diagnosis for the same person while correcting for agreement that would occur by chance alone. Specificity and sensitivity values were also assessed to indicate whether the EDA-5 and EDA-5 apps could correctly predict whether an individual met diagnostic criteria for anorexia nervosa, bulimia nervosa, binge-eating disorder, other specified feeding

or eating disorder (OSFED), and unspecified feeding or eating disorder (UFED) using either clinical diagnosis or EDA-5 diagnosis as the reference.

Reliability Test-retest reliability for the EDA-5 paper-and-pencil version over an average of 9 days was excellent across diagnoses ($\kappa = 0.87$).

Validity Concurrent validity between the EDA-5 paper-and-pencil version and the EDE ranged from good ($\kappa = 0.65$ for OSFED and UFED) to excellent ($\kappa = 0.90$ for BED), with an average of $\kappa = 0.74$ across diagnoses. Most diagnostic discrepancies between the EDA-5 and EDE were associated with differences in how low body weight was determined in each assessment tool. For example, a body mass index of 18.5 or less at any time in the past 3 months is coded as meeting the low body weight criterion for AN in the EDA-5, whereas the EDE considers only current weight and height in the past month.

Concurrent validity between the EDA-5 app and an unstructured clinical interview was 87.3%. Kappa coefficients ranged from fair ($\kappa = 0.56$ for OSFED and UFED) to excellent ($\kappa = 0.94$ for bulimia nervosa) with an average of $\kappa = 0.83$. Concurrent validity between the EDA-5 app and the Eating Disorder Diagnostic Scale (EDDS; [45]) varied across diagnosis and ranged from poor ($\kappa = 0.27$ when no eating disorder diagnosis is made) to excellent ($\kappa = 0.77$ for AN). Diagnostic discrepancies between the EDDS and the EDA-5 app generally occurred because individuals denied functional impairment on the EDDS, but not on the EDA-5 app. Thus, it is possible that the EDA-5 app may be a more sensitive measure of clinical impairment secondary to eating disorders compared to the EDDS.

Sensitivity was calculated to identify the proportion of participants who met criteria for an eating disorder as determined by the EDE who were also identified as having that same eating disorder using the EDA-5 paper-and-pencil version. Sensitivity values for the paper-and-pencil version ranged from 0.65 for OSFED/UFED to 1.00 for anorexia nervosa and binge-eating disorder. Sensitivity values for the electronic version of the EDA-5 range from 0.73 for OSFED/UFED to 0.96 for bulimia nervosa. Specificity was calculated to identify the proportion of participants who did not meet criteria for an eating disorder when assessed with the EDE who also did not meet criteria for that same diagnosis when using the EDA-5 paper-and-pencil version. Specificity values for the paper-and-pencil version ranged from 0.83 for AN to 0.98 for BED. Specificity values for the EDA-5 app ranged from 0.90 for OSFED/UFED to 1.00 for AN.

Finally, EDA-5 diagnoses showed evidence for convergent and discriminant validity because they were associated with appropriate variation in self-report measures of eating-disorder behaviors (e.g., those diagnosed with bulimia

nervosa or binge-eating disorder had higher scores on the Eating Pathology Symptoms Inventory (EPSI [46••]) binge-eating scale and lower scores on the restricting scale compared to individuals with AN).

Strengths and Limitations The EDA-5 takes less time to administer than the EDE, although the amount of time saved differed across administration sites (the amount of time saved ranged from approximately 8 to 14 min). Interviewers who administered the EDA-5 in the scale development and validation study reported that the EDA-5 was less “complex” and more focused on important symptomatology compared to the EDE.

Another useful feature of the EDA-5 is that it can be administered via paper-and-pencil or electronically, with the electronic version further reducing administration time compared to the paper-and-pencil version. The EDA-5 app saved interviewers an average of 5 min compared to the EDA-5 paper-and-pencil version. Because the EDA-5 was developed using eating-disorder tertiary care sites, future research is needed to determine the reliability and validity of the EDA-5 for diagnosing eating disorders in community samples and in general outpatient treatment settings. In addition, the EDA-5 included few men in the development and validation sample; thus, the psychometric properties of the EDA-5 in boys and men are unclear.

Eating Disorder Questionnaire-Online

Purpose and Content The Eating Disorder Questionnaire-Online (EDQ-O) [47••] is a brief self-report measure originally written in Dutch (English translation provided in the publication) that was designed to assign eating-disorder diagnoses consistent with the *Diagnostic and Statistical Manual of Mental Disorders—Fourth Edition, Text Revision (DSM-IV-TR)* [48]. This measure was developed because there were no previous online self-report assessments that assigned DSM-IV-TR eating-disorder diagnoses (note: the Eating Disorder Examination-Questionnaire [49], a widely used self-report measure of eating pathology, did not allow for DSM binge-eating-disorder diagnosis at the time the EDQ-O was developed).

The EDQ-O is administered completely online and includes questions that correspond to DSM-IV-TR diagnostic criteria for anorexia nervosa, bulimia nervosa, binge-eating disorder, and eating disorder not otherwise specified (EDNOS). The EDQ-O uses a skip-logic algorithm to skip-out any of the questions that do not apply to the individual completing the assessment (e.g., if an individual is not at a low body weight, anorexia nervosa questions are not administered). The EDQ-O is a potentially useful diagnostic measure that individuals can complete remotely (in lieu of in-person

clinical interviews). The measure was validated according to the Longitudinal, Expert, and All Data (LEAD) standard [50].

Appropriate Uses The EDQ-O is appropriate for use in assigning DSM-IV-TR eating-disorder diagnoses in adults, although it was developed in a sample ranging in age from 16 to 60 years. There is no child or adolescent version. It is appropriate to use as an online diagnostic tool in addition to, or in place of, clinical interviews (e.g., if setting or time constraints do not allow for a semistructured clinical interview, the EDQ-O may be appropriate).

Scale Development The EDQ-O was developed in a Dutch sample of 134 individuals (88% women) who were receiving treatment at an eating-disorder specialty clinic. Participants completed the EDQ-O remotely prior to completing an in-person clinical diagnostic assessment. Development of anorexia nervosa and bulimia nervosa diagnostic questions for the EDQ-O was based on the Mini-International Neuropsychiatric Interview (MINI) [2, 51]; binge-eating-disorder diagnostic questions were adapted from DSM-IV-TR research criteria [48]. Individuals who endorsed eating-disorder symptoms but who did not meet DSM-IV-TR criteria for anorexia nervosa, bulimia nervosa, or binge-eating disorder were assigned a diagnosis of eating disorder not otherwise specified.

Reliability No reliability data are available at this time.

Validity Concurrent validity was tested by comparing diagnoses derived from the EDQ-O to diagnoses derived from clinical interviews according to the LEAD standard [50]; agreement between the EDQ-O and clinical interview was estimated using the area under the receiver operating characteristic (ROC) curve. The EDQ-O demonstrated acceptable classification accuracy (AUC range = 0.72–0.83) and agreement (accuracy range = 79–93%) with LEAD standard diagnoses for all eating-disorder diagnoses. Positive predictive values and negative predictive values were good for all diagnoses except bulimia nervosa (positive predictive value = 0.50) and EDNOS (positive predictive value = 0.75). The authors did not evaluate convergent or discriminant validity.

Strengths and Limitations A strength of the EDQ-O is its brevity. Indeed, completion time is estimated to be 5 min. This measure is available completely online, which may facilitate screening/diagnosis (if a clinical interview is not possible) and interventions that are web-based. The EDQ-O is available in English; however, it was written and validated in Dutch-speaking participants; thus, the psychometric properties in English-speaking samples have not been established. Although the EDQ-O is based on DSM-IV-TR diagnostic criteria, the translation to English is not word for word. Certain questions feature colloquialisms that may not be

familiar to English-speaking persons (e.g., to describe an example of a binge-eating episode, respondents are provided with examples such as: “...one or more bags of chips along with a chocolate bar or a *pack of magnums*...”) or verbiage that is not consistent with common phrasing in English (e.g., “Can you give a description of the foods ... *you often take* during a binge?”). Although examples are provided as benchmarks for the amount of food consumed during an objective binge-eating episode, no information is provided about inappropriate compensatory behaviors; thus, questions about what behaviors are considered to represent “purging behaviors” are left open to interpretation.

Another concern is that the EDQ-O does not allow for DSM-5 [43] eating-disorder diagnoses to be assigned; this is of particular concern for anorexia nervosa because the EDQ-O includes anorexia nervosa criterion D (amenorrhea), which likely excludes a number of women who exhibit all other symptoms of anorexia nervosa, but who would not be diagnosed as having anorexia nervosa based on the EDQ-O’s skip-logic algorithm. These individuals would be assigned a diagnosis of an eating disorder not otherwise specified, with no additional information provided. Finally, it is important to note that information about the reliability of the EDQ-O is currently unavailable.

Eating Pathology Symptoms Inventory

Purpose and Content The EPSI [46] is a 45-item measure that was designed to assess the psychopathology of eating disorders. The EPSI contains eight scales that assess *body dissatisfaction* (dissatisfaction with body weight and/or shape), *binge eating* (ingestion of large amounts of food and accompanying cognitive symptoms), *cognitive restraint* (cognitive efforts to limit or avoid eating, whether or not such attempts are successful), *purging* (self-induced vomiting, laxative use, diuretic use, and diet pill use), *excessive exercise* (physical exercise that is intense and/or compulsive), *restricting* (concrete efforts to avoid or reduce food consumption), *muscle building* (desire for increased muscularity and muscle building supplement use), and *negative attitudes toward obesity* (negative attitudes toward individuals who are overweight or obese). The EPSI is free for noncommercial clinical or research use. A copy of the EPSI can be obtained and downloaded from <https://psych.ku.edu/kelsie-t-forbush> or from the NIH PhenX Toolkit at: <https://www.phenxtoolkit.org/index.php?pageLink=browse.si.all&nimh=true>. The EPSI can also be administered within the Recovery Record, Inc. mobile phone application for free (for patient users) or for a fee (for clinician users).

Appropriate Uses The EPSI can be administered as a full measure or a subset of scales can be administered, based on the administrator’s preference. The EPSI was developed in

samples as young as 14 years of age and has a Flesch-Kincaid reading level of sixth grade; thus, use of the EPSI in younger children and teenagers (below age 14 years) is not currently recommended. The EPSI can be used across a wide range of contexts, including community members, college students, and patient groups, and across the weight spectrum. Although the EPSI generally performs well across sexes, the muscle building scale performs less well in women (see sections below). Thus, caution should be used when interpreting muscle building scores in female patients or in groups that are mostly comprised of women. Normative data in college men and women can be found in Forbush et al. [52]. Patient norms (including normative data for persons with anorexia nervosa or bulimia nervosa can be obtained from Forbush et al. ([46••] in Table 5).

Scale Development The scale development process involved creating an initial item pool of 160 items to assess 20 theoretical dimensions of eating disorders (including, but not limited to, all of the symptoms in the DSM [43]). The EPSI item pool was developed based on extant theoretical models of eating disorders, expert input, and from a review of relevant research literature. The initial item pool was administered to 433 college students (62.59% women) and 407 (47.4% women) community volunteers. Next, factor analysis (both exploratory and confirmatory) was used to identify the underlying structure of the measure to develop preliminary scales. The authors also conducted invariance tests to determine whether the structure of the EPSI was comparable in men versus women and between persons of normal weight or with overweight/obesity. Based on the results of empirical analyses of the preliminary item pool, new items were written to better assess certain constructs and the revised measure was administered to psychiatric outpatients ($N = 303$; 65.58% women) and patients with an eating disorder ($N = 158$; 94.3% women). Additional exploratory and confirmatory factor analyses were used to identify a final structure of the measure, which were used to form the eight EPSI scales.

Finally, the EPSI was administered to college students ($N = 227$; 58.15 women) to measure test-retest reliability over a 2-week period.

Reliability Internal consistency reliability for the EPSI was generally good to excellent across a variety of populations, with median coefficient alpha values ranging from 0.84 to 0.89 [46]. One important caveat was that the muscle building scale had lower internal consistency in women (ranging from 0.37 to 0.54). As we describe in the section on validity (see below), the muscle building scale performed well in men but may have lowered reliability and validity in women. Test-retest reliability for the EPSI was evaluated in a sample of 227 college students tested twice over a 2-week period [46]. Results suggested that EPSI scales were stable over time with

most scale retest reliabilities > 0.70 . Recent research by Forbush et al. [53] found that EPSI scales were reliable over a 1-month period in both men and women. Moreover, the EPSI was one of the few measures (out of 10 tested) that provided a reliable measure of binge-eating and inappropriate compensatory behaviors, and was equally reliable in men and women [53]. On the other hand, it is important to note that test-retest reliability for the muscle building scale, which was excellent in men ($ICC = 0.79$), was poor in women ($ICC = 0.47$).

Validity Forbush et al. [46••] reported that one of the reasons that they developed the EPSI was to address issues with poor discriminant validity (moderate to high correlations among constructs that are not supposed to relate strongly) that had been identified in existing eating-disorder self-report measures. For example, past research has shown that some measures of eating-disorder symptoms (e.g., the EDI bulimia scale) correlated more strongly with measures of depression than with other similar eating-disorder symptoms (e.g., the Multi-factorial Assessment of Eating Disorder Symptoms Purging scale) [54]. Research across a range of samples found that EPSI scales showed evidence for excellent discriminant validity from mood- and anxiety-related content [46, 52]. EPSI scales also correlated moderately to strongly with related scales from the EDI-3, Eating Disorder Examination-Questionnaire, Body Shape Questionnaire, Male Body Attitudes Scale, and several others, providing evidence for convergent validity [46, 52].

EPSI scales show evidence for strong construct validity in men and women; however, results from multiple group analysis (which tests whether items that form a scale can be interpreted in the same way between groups) showed that the muscle building scale did not perform as well in women [46]. This is likely because women were generally less concerned with achieving a more muscular body than with reducing body fat and, as a result, had relatively low rates of steroid and protein supplement use.

One important feature of the EPSI was that scales have shown evidence for strong criterion-related validity by distinguishing among patients with eating disorders from noneating-disorder controls and among various types of eating-disorder diagnoses [42, 46]. In particular, the restricting scale differentiates patients with anorexia nervosa from noneating-disorder controls and patients with bulimia nervosa or binge eating disorder, whereas the binge-eating scale distinguished those with bulimia nervosa or binge-eating disorder from patients with anorexia nervosa and noneating-disorder controls. Finally, although content validity (the extent to which a measure represents all domains of the construct) is not often discussed in the context of eating-disorders assessment, the EPSI appears to be one of the most thorough and comprehensive measures of eating pathology, which may make it a

useful measure for routine clinical care and in research that seeks to answer substantive questions about the full range of eating-disorder behaviors.

Strengths and Limitations Strengths of the scale development process for the EPSI included the incorporation of both theory and statistical methods to develop the final version from the initial item pool, inclusion of sufficient numbers of men (an often excluded population in eating-disorder research studies) and persons with overweight and obesity, and administration of the measure to several large samples, including patient and nonpatient groups. The major strengths of the EPSI include the strong psychometric properties across numerous populations (including in nonpatients, patients, in men and women, and across the weight spectrum [55]). The eight-factor structure of the EPSI has been replicated in several studies, including in a sample of native Chinese speakers [56]. The EPSI has been validated in both online and paper-pencil formats. Overall, the EPSI shows substantial evidence for reliability and validity, which makes it a useful self-report measure for clinical and research use. Compared to traditional “gold-standard” measures of eating-disorder behaviors (e.g., the EDI-3 and Eating Disorder Examination-Questionnaire), the EPSI has similar (or better) psychometric properties and a more replicable factor structure. The EPSI also appears to cover a full range of eating-disorder behaviors, including behaviors that are relevant to eating pathology as it presents in men (e.g., muscle building).

Limitations of the EPSI include evidence that the muscle building scale is less reliable and valid in women. Thus, clinicians and researchers who administer the muscle building scale to female patients or participants should use caution when interpreting that specific scale with women. Another limitation is that the majority of research on the psychometric properties and clinical utility of the EPSI have been published by Forbush and colleagues, and additional studies are needed to replicate findings. Despite limitations, the EPSI appears to be a promising measure of eating-disorder symptoms that can be used across a wide range of research and clinical settings.

The Interactive, Graphical Assessment Tool

Purpose and Content The Interactive, Graphical Assessment Tool (IGAT) [57••] is a computer-based, self-report measure that assesses the frequency of disordered-eating behaviors over the past 12 weeks, including both subjective and objective binge-eating episodes, purging behaviors (e.g., self-induced vomiting; laxative, diuretic, enema, and suppository misuse; and misuse of insulin or other medications), fasting, and exercise undertaken in response to dissatisfaction with body shape or weight. The IGAT also assesses weekly body weight fluctuations (in pounds) and subjective, user-defined stress levels. A demonstration version of the IGAT can be

found here: <http://undeatingbehaviors.wixsite.com/uwyoeatingbehaviors/interactive-graphical-assessment-tool>.

When clients or participants begin the IGAT, they are presented with a calendar that displays the past 12 weeks. Individuals are asked to note “landmark” personal events (e.g., birthdays) from an events menu. “Landmark” public events (e.g., holidays) are prepopulated into the calendar. Clients or participants provide information about the presence of disordered-eating behaviors via a checklist and report their highest and lowest body weights over the past 12 weeks. Next, individuals complete a guided tutorial before indicating their average body weight in pounds for each of the past 12 weeks. After the tutorial, the respondent provides the frequencies (in days and number of episodes) that he or she engaged in disordered-eating behaviors each week. Finally, participants are asked to indicate their average stress level for each of the past 12 weeks on a scale of “low” to “high.”

The IGAT provides a graphical interface for respondents to input data. Specifically, respondents move dots along a line graph to report number of disordered-eating behavior days for each disordered-eating behavior endorsed, and average body weight and stress levels for each of the past 12 weeks. The number of disordered-eating behavior episodes per week is manually entered by the respondent in boxes below the corresponding week. If a client or participant reports fewer disordered-eating episodes than reported disordered-eating days, the IGAT prompts the participant to double-check responses. The IGAT contains separate line graphs for separate symptoms and line graphs are differentiated by tabs; for example, one line graph assesses purging frequencies while another assesses binge-eating episode frequencies. Throughout the assessment, the IGAT displays a calendar of the past 12 weeks marked with landmark events below the line graph. The IGAT also displays a detailed calendar that corresponds to the week that is being assessed. The purpose of calendar presentation is to facilitate respondents’ ability for accurate retrospective recall of when disordered-eating behaviors occurred.

Appropriate Uses The IGAT is appropriate for use in persons seeking or receiving eating-disorder treatment, particularly when time constraints limit the feasibility of administration of semistructured diagnostic interviews. The authors reported that the IGAT takes approximately 8–14 min to complete, which is shorter than most diagnostic interviews. The IGAT may be particularly useful when an individual first presents to treatment because the IGAT provides a “snapshot” of disordered eating over the past 3 months, which may facilitate diagnosis. The IGAT may also be useful for assessing disordered-eating symptom trajectories over the course of treatment or in research settings when retrospective (vs. prospective) reports of disordered-eating behaviors are of interest. The IGAT was developed in a sample of men and women

(82.3% women; aged 18 to 72 years) who presented with a variety of eating-disorder diagnoses. Information regarding the psychometric properties of the IGAT in persons under 18 years of age, in noneating-disorder samples, or in all-male samples is unavailable.

Scale Development The IGAT was designed to facilitate more accurate retrospective reporting of eating-disorder behaviors by providing concrete beginning and ending dates for the assessment period (i.e., bounding), asking respondents to list “landmark” personal dates on a calendar, providing a tutorial, and giving frequent visual cues (e.g., graphs, calendar presentation). The IGAT was pilot-tested in small focus groups comprised of college-aged women and persons with eating disorders; each focus group contained approximately four to six members (K. De Young, personal communication, May 9, 2017). No questions were modified or removed from the initial item pool prior to the development of the finalized version of the IGAT. Test-retest reliability and convergent and criterion-related validity were tested in a sample of persons with eating disorders ($N = 113$; 82.3% women).

Reliability Six-week test-retest reliability was computed using ICC. The IGAT was administered at baseline (IGAT_B), 6 weeks (IGAT₆), and 12 weeks (IGAT₁₂). The authors computed test-retest reliability by comparing weeks reported upon twice across IGAT assessments through overlapping timeframes: (1) weeks 6 through 12 of IGAT_B and weeks 1 through 6 of IGAT₆; and (2) weeks 6 through 12 of IGAT₆ and weeks 1 through 6 of IGAT₁₂. For simplicity, we report the range of ICC across assessment intervals. Six-week test-retest reliability coefficients were excellent for body weight (ICCs = 0.993–0.998). Test-retest reliability was fair to good for binge-eating days (ICCs = 0.465–0.647) and poor to fair for number of binge-eating episodes (ICCs = 0.190–0.566). Retest reliability for purging was good to excellent for purging days (ICCs = 0.687–0.829) and poor to excellent for purging episodes (ICCs = 0.309–0.753). Test-retest reliability was fair to excellent for fasting (ICCs = 0.493–0.863) and exercise (ICCs = 0.509–0.810). Finally, test-retest for stress was poor to fair (ICCs = 0.251–0.466).

Validity Convergent validity was evaluated by correlating IGAT scales with the Weekly Self-Monitoring Questionnaire (WSMQ) [57••], which was created by the IGAT authors to assess disordered-eating symptom frequency and body weight in pounds on a weekly basis. Convergent validity of IGAT body weight with WSMQ body weight was excellent (ICCs = 0.979–0.998). Convergent validity was fair to good for binge-eating days (ICCs = 0.444–0.690) and fair to excellent for binge-eating episodes (ICCs = 0.421–0.776). For purging, convergent validity was good to excellent for purging days (ICCs = 0.726–0.894) and fair to excellent for purging

episodes (ICCs = 0.447–0.811). Lastly, convergent validity was fair to excellent for fasting (ICCs = 0.440–0.876) and fair to good for exercise (ICCs = 0.414–0.716).

Additionally, convergent validity of IGAT binge-eating days and episodes and purging episodes was tested via correlations with similar scales of the EDE-Q [49]. The concurrent validity of IGAT and EDE-Q was poor to excellent for binge-eating days (ICCs = 0.323–0.848) and poor to good for binge-eating episodes (ICCs = 0.323–0.709). Convergent validity of IGAT and EDE-Q purging episodes was good to excellent (ICCs = 0.641–0.788). Convergent validity of the IGAT stress scale was tested via correlations with the negative affect scale from the positive and negative affect schedule [58] and ranged from weak to moderate (Spearman’s rho = 0.173–0.516).

Finally, receiver operating characteristic curve analyses were conducted to test the sensitivity and specificity of IGAT’s ability to identify disordered-eating behaviors (binge eating, purging, fasting, exercise) at diagnostic threshold levels according to the DSM-5 frequency cutoff (e.g., eating-disorder behaviors that occurred at least once per week on average over the past 3 months), using the WSMQ as the criterion. Sensitivity values ranged from 0.57 for exercise to 0.81 for purging. Specificity values ranged from 0.82 for binge eating to 1.0 for purging. The IGAT demonstrated acceptable classification accuracy (area under the curve [AUC] = 0.81–0.91) for disordered-eating behaviors.

Strengths and Limitations The IGAT provides a user-friendly, online platform to assess disordered-eating behaviors over the past 12 weeks. The developers of the IGAT used empirically supported methods to increase the accuracy of retrospective symptom reporting, such as line graphs and a detailed calendar with landmark events, which may decrease participant burden by facilitating accurate reporting of past eating-disorder behaviors. The IGAT also provides a tutorial, which is helpful in promoting effective use of the tool and minimizing participant confusion.

The IGAT shows evidence for excellent test-retest reliability for exercise, fasting, and body weight. The IGAT also demonstrated evidence for good convergent validity for disordered-eating behaviors and body weight. On the other hand, it is important to note that test-retest reliability coefficients were lower for binge-eating and purging episodes and stress, and convergent validity for binge-eating days and episodes with the EDE-Q ranged from poor to excellent. Lower test-retest reliability coefficients for IGAT binge-eating and purging episodes suggested that the IGAT might be less sensitive to capturing true changes (or have greater measurement error) for assessing binge eating and purging. The authors explain that lower convergent validity of IGAT binge-eating days and episodes with the EDE-Q could be explained by EDE-Q’s tendency to *underestimate* binge-eating frequencies. However, other work has indicated that binge-eating and

purging behaviors often are less reliable over time across the majority of existing eating-disorder questionnaires [53], so this issue is not specific to the IGAT. Another limitation is that the IGAT assesses body weight, but not height—therefore, BMI cannot be computed. BMI is particularly important for differential diagnosis; thus, the IGAT should be supplemented with information about objective or self-reported height to facilitate diagnosis.

Finally, although the purpose of the IGAT was to provide an accurate retrospective assessment of eating-disorder symptoms, there are no studies that tested whether the IGAT provides more valid retrospective recall of eating-disorder behaviors compared to traditional eating-disorder measures, such as the EDE-Q. Despite these limitations, the IGAT represents a novel eating-disorders assessment tool that can be administered via the Internet; is low burden to the participant due to relatively short completion times, as well as features that facilitate participants' memory (e.g., graphs, calendars, landmark events); and provides data regarding disordered-eating behavior frequency and body weight fluctuations.

Discussion

The purpose of this paper was to review novel and innovative multidimensional eating-disorder assessments that were developed within the past 5 years for use in clinical and/or research settings. The measures we reviewed included self-report measures (CR-EAT, ED-15, EDQ-O, EPSI, and IGAT) and one structured clinical interview (EDA-5). One important feature was that all measures included at least one clinical sample in the scale development or validation process. Including clinical samples is important because it increases the likelihood that these new measures will be appropriate for use in the populations to which they will be applied. Thus, there is lower concern that the assessments we reviewed will perform differently in clinical samples compared to the original reports.

Another innovative aspect of newly developed eating-disorder measures is that most measures (except for the ED-15) are available for participants to complete online. In the era of Internet interventions and e-mental health services, the ability to complete measures online reduces respondent burden because they do not need to travel to a clinic or research laboratory to complete assessments. In addition, online assessment reduces data entry burden for staff and may help with interpretation, if results and norms are provided to the researcher or clinician electronically. Both the EPSI, which is available within the Recovery Record, Inc. mobile phone app, and the IGAT have attractive user-friendly delivery platforms. The EPSI provides feedback directly to the respondent about improvements and worsening of symptoms, which may improve the user experience and facilitate interest in completing

repeated measures. The IGAT provides helpful tutorials, which reduce the likelihood of respondent mistakes, and a calendar for the respondent to include relevant personal events to lessen the impact of retrospective recall errors.

An exciting aspect of the measures we reviewed was that certain assessments introduced novel constructs or diagnoses that were not a part of other existing multidimensional eating-disorder tools. For example, the EPSI *Negative Attitudes toward Obesity* scale assesses social attitudes toward others' bodies that may impact treatment outcomes via internalized weight stigma (viz., if a client has negative views about other peoples' overweight or obese bodies, this may affect his or her own attitude about restoring or maintaining his or her own body weight). The EPSI also includes *Muscle Building*, which is an important component of eating-disorder psychopathology as it presents in men [59]. Thus, the EPSI may provide a more content-valid, comprehensive understanding of the respondent's disordered-eating behaviors. The EDA-5 represents an important new diagnostic assessment that includes avoidant and restrictive food intake disorder (ARFID), which is not included in traditional eating-disorder diagnostic assessments, such as the EDE. The EDA-5, therefore, may be particularly helpful in contexts in which differential diagnosis between anorexia nervosa and ARFID is needed.

Finally, certain measures provided information that could be particularly helpful for clinical interpretation purposes. The ED-15 is notable as a week-to-week outcome measure and may be useful in treatment settings or in clinical research trials. The ED-15 provided community-based centile scores, which can be used to gauge when clients' disordered-eating behaviors have returned to more normative levels over the course of an intervention. In summary, the strengths of novel eating-disorder assessments include immediate applicability to clinical populations, expanded range of content, availability within the online environment, and week-to-week retrospective measurement of disordered eating for use in research or clinical settings (ED-15 and IGAT).

Limitations of Reviewed Measures

Despite the strengths of the new eating-disorder assessments we reviewed, there are important limitations. All of the measures we reviewed were tested in primarily Caucasian samples, despite research showing that eating disorders occur in all ethnicities and races [60]. Another key limitation is that, with the exception of the EPSI, most new measures were developed and validated in mostly female samples (see Table 2), with 71.4–100% of samples comprised of women. Thus, it is unclear whether the psychometric properties of certain newly developed measures vary between sexes. Research on the EPSI demonstrates the importance of testing eating-disorder measures in both men and women. For example, the EPSI muscle building scale showed evidence for

Table 2 Description of self-report assessments for eating disorders

Measure	Purpose	Administration	Scales	No. of items	Population	Sample size	% Female
Clinical and Research Inventory for Eating Disorders (CR-EAT) [30••]	Used to assess the symptomatology of eating disorders for clinical and research purposes	Online self-report	Weight preoccupation, body embarrassment, restrained eating behavior, societal expectations of weight and shape, harmful weight regulation, mood dysregulation, affect-regulatory eating, self-esteem, concerns about negative evaluation, perfectionism: personal expectations, perfectionism: familial expectations	63	Nonclinical sample, nonclinical paper-and-pencil sample, clinical sample	1406, 220, 61	74.3, 71.4, 100
ED-15 [38••]	A session-by-session self-report measure of eating disorders attitudes and behaviors over a period of 1 week	Paper-and-pencil self-report	Weight and shape concerns, eating concerns	15	Nonclinical adults, adults with self-reported eating disorders, formally diagnosed adults receiving CBT	531, 63, 33	82.4, 100, 100
Eating Disorder Assessment for DSM-5 (EDA-5) [42••]	A DSM-5 diagnostic tool intended to be used primarily in research settings by assessors with limited training. The EDA-5 only assesses feeding and eating disorder criteria, not other related psychopathology	Semistructured interview, paper-and-pencil or online	AN, BN, BED, ARFID, Pica, rumination disorder, OSFED, USFED	Number of items administered varies based on individual's responses	Treatment-seeking adolescents and adults	66, 72	89.1, 94.4
Eating Disorder Questionnaire-Online (EDQ-O) [47••]	A strictly online DSM-IV-TR diagnostic tool for AN, BN, BED, and EDNOS	Online self-report	AN, BN, BED, EDNOS	Number of items administered varies based on individual's responses (37 total)	Adults receiving treatment at an eating disorder specialty clinic	134	88
Eating Pathology Symptoms Inventory (EPSI) [46••]	Used to assess the psychopathology of eating disorders	Paper-and-pencil or online self-report within the Recovery Record, Inc. mobile phone application	Body dissatisfaction, binge eating, cognitive restraint, purging, excessive exercise, restricting, muscle building, negative attitudes toward obesity	45	College students, community sample, psychiatric outpatients, patients with an eating disorder, college students	433, 407, 303, 158, 227	62.59, 47.4, 65.58, 94.3, 58.15
Interactive, Graphical Assessment Tool (IGAT) [49]	Designed to assess the frequency and severity of eating disorder behaviors over a period of 12 weeks	Online self-report	Body weight, binge eating, purging, fasting, exercise, stress	Number of items administered varies based on individual's responses	Persons with eating disorders	113	82.3

AN anorexia nervosa, ARFID avoidant and restrictive food intake disorder, BN bulimia nervosa, BED binge-eating disorder, EDNOS eating disorder not otherwise specified, OSFED other specified feeding and eating disorder

excellent reliability and validity in men, but not in women, which highlights the importance of testing constructs across specific demographic groups. Finally, with the exception of the EPSI, no studies tested sufficient numbers of persons with overweight or obesity to determine whether there was differential reliability or validity across the weight spectrum, which is important given the increasing prevalence of overweight and obesity in Western society [61] and because eating disorders occur across the weight spectrum.

The most salient limitation was poor adherence to recommended scale development and testing procedures. Several well-respected authors [28, 41, 62] have provided consistent recommendations for scale development practices; yet, in many cases, these recommendations were not followed. For example, scale development experts recommend that (1) initial item pools should be developed using rational methods (theory) and (2) empirical methods should be used to remove suboptimal items from the initial item pool (statistics). The inclusion of both rational and empirical methods typically results in an iterative process that includes several rounds of item writing, questionnaire administration (and re-administration), testing in large samples, and statistical analysis to identify items for elimination prior to finalizing the item pool. Moreover, appropriate scale development procedures highlight the importance of removing items from the initial pool if they show evidence for poor psychometric properties, including (but not limited to): (1) factor loadings < 0.40 on primary factors or cross-loadings > 0.30 on secondary factors in exploratory factor analysis; and (2) differential item function between groups. Moreover, scales should be revised if they show (1) low correlations with constructs that should be theoretically related (e.g., body dissatisfaction should be correlated with other scales of body dissatisfaction) and/or (2) high correlations with constructs that are theoretically distinct (e.g., a binge-eating item should not correlate strongly with a measure of negative affect). Following these recommendations is important because it helps to prevent issues with low internal consistency or test-retest reliability and ensure good construct validity.

The EPSI was the only assessment that underwent an iterative process that included multiple stages of item writing and testing. Only three measures (CR-EAT, ED-15, and EPSI) used principle components analysis or factor analysis to eliminate poorly performing items from the initial item pool. Other measures did not appear to use any empirically based methods for eliminating poorly performing questions from the initial item pool, which may have led to weaker construct validity (see description below). Moreover, the EDQ-O did not report any data on reliability in the development or validation of the instrument; the final measure only reported criterion-related validity using ROC curve analysis.

Consistent with poor adherence to “best practices” for scale development, we noted several errors in the usage of common psychometric terminology that resulted in incomplete tests of

construct validity. Several authors tested convergent or criterion-related validity but did not assess discriminant validity. Other authors mislabeled analyses that were focused on distinguishing groups (criterion-related validity) as discriminant validity. Both classic definitions of construct validity [63] and modern scale development terminology [28] describe construct validity as comprising convergent and discriminant validity. Discriminant validity is supported when there are low to moderate correlations among constructs that are thought to be unrelated, whereas convergent validity is supported by strong correlations among theoretically related constructs (see Table 1 for additional definitions).

The lack of standard psychometric terminology has also led to poor testing or reporting procedures, given that many measures did not adequately test construct validity—which many assessment scholars consider to be the most important form of validity [41]. Many authors tested convergent validity but did not assess discriminant validity, which was poor for several measures. In some instances, we found evidence that the correlation between two similar constructs (e.g., two binge-eating measures) was *lower* than for two measures of *different* constructs (e.g., binge eating and body dissatisfaction). We contend that correlations among related constructs should be stronger than correlations of unrelated constructs in order for a measure to have maximal clinical utility and validity. Related to this issue is that for certain eating-disorder scales, we found higher correlations with depression-related scales than with other eating-disorder symptom scales. Logically, two measures of eating-disorder symptoms should have higher correlations with each other than with mood and anxiety-related constructs. Otherwise, in treatment settings, it is unclear whether the measure is testing changes in eating-related attitudes or depression/anxiety over the course of treatment. Issues with discriminant validity from anxiety and depression could be addressed through adhering to item-writing procedures that recommend de-emphasizing wording that inadvertently taps negative affect (e.g., terms such as “distressed” and “worried”). This may be why discriminant validity for some of the measures that we described in our review was lower than ideal.

Future Directions

Future directions include the need for expanded testing and validation for all measures. The psychometric properties for the majority of assessments (other than the EPSI) have not been tested outside of the initial publication. Moreover, the authors of the EPSI conducted the majority of testing on its reliability and validity, and the EPSI’s factor structure and reliability need to be confirmed in independent replications. Given the development of novel methods for tracking physical activity, social attitudes (such as modified versions of the Implicit Attitudes Test), and the ability to directly observe

behaviors (such as body checking as measured in an experimental design or laboratory feeding studies), the measures we described in our review could be further tested for criterion-related validity using objective methods.

Another issue that cuts across the measures we described is that few measures provided information to help clinicians interpret clients' scores. In other areas of psychological assessment (e.g., the MMPI-2 and PAI), there are detailed manuals that provide norm-referenced information to help clinicians understand how their client's scores compare to those in the general community or in patient samples. Additional information to explain what scores mean, how they relate to normative data, and what expected patterns of change should be in various treatment settings would provide a much more solid basis for clinicians to assess their client's progress in treatment (or lack thereof). Similarly, although most measures included pilot testing and feedback from experts, to our knowledge, no prior work has asked for clinician's perspectives on the utility of various traditional "gold-standard" and novel measures, which may be important for ensuring that assessments are meeting the needs of practitioners.

Few novel assessments were tested across the full range of populations that experience eating disorders. Other than the EPSI, there are few-to-no tests of the psychometric properties of the newly developed measures in men or boys. None of the measures we reviewed were tested in sufficient numbers of ethnic or racial minorities to determine whether the scales perform equally well across groups. Given the availability of statistical methods for identifying poor scales and items (such as multiple-group analysis within structural equation modeling and item response theory), these could be powerful tools for developing instruments that are maximally reliable across the range of people with eating disorders. In addition, all the measures we described were appropriate for use in adult populations, but no measures were appropriate for use in younger populations (e.g., in youth age 10 and above) with or without eating disorders. The development of measures for use in children and young adolescents would be valuable for studies designed to identify youth at risk for the development of eating disorders, as well as for use in research and treatment settings, particularly considering the number of individuals for whom eating-disorder onset is in childhood or early adolescence.

Finally, there was substantial diversity in the separate constructs that were assessed in the measures we reviewed. Although we do not view this as a limitation of the existing literature on eating-disorder assessment, it provides an interesting opportunity for future research. If researchers create initial item pools that are based on theoretical models, then scale development provides an exciting opportunity to test the specificity and breadth of the construct. For example, some measures focused on DSM symptoms, others were broader in coverage but still focused on core eating-disorder symptoms,

whereas other measures included core eating-disorder symptoms and constructs implicated in etiological, maintenance, and treatment models (e.g., affect regulation and stress). Using appropriate scale development methods and statistical analyses for testing convergent and discriminant validity, therefore, may allow the field to better "carve nature at its joints" through an improved ability to understand, refine, and measure the construct of eating psychopathology.

Conclusions

We identified six novel eating-disorder measures that have been published since 2013. These included five self-report measures and one structured clinical interview. All measures (except for one) are available in an online format, which provides many advantages for administration and scoring. Other advantages were that all measures we reviewed included clinical samples in their initial validation, which increases the likelihood that the assessments will translate into useful tools for clinical practice or in clinical research settings. Additional notable features of the novel assessments we reviewed included the following: improved content validity to assess the full range of eating pathology, new formats that were designed to help reduce retrospective recall bias, and assessment timeframes that enable tracking of week-to-week changes for psychotherapy practice or clinical trials.

Despite the numerous strengths of novel approaches to eating-disorder assessment, there are still limitations that should be considered for future research. Most importantly, certain measures did not adequately report validity data (particularly discriminant validity) or had poor discriminant validity from theoretically nonrelated constructs. To the extent that measures have poor discriminant validity, they may not provide optimal assessments of the intended constructs, which may hamper the ability to make meaningful clinical recommendations and research conclusions. Thus, future research may benefit from a greater focus on careful testing and validation, using recommended scale development procedures, to ensure that researchers and clinicians are able to assess clients and participants using psychometrically sound tools. Indeed, assessment forms the foundation for understanding clinical changes over time, informs our understanding of the psychopathology of eating disorders, and is critical for studies that require valid phenotypes to better link to genetic and biologically informed research.

Compliance with Ethical Standards

Conflict of Interest Kelsie T. Forbush receives commercially funded research grant monies to include a new version of the Eating Pathology Symptoms Inventory within the Recovery Record, Inc. mobile phone application. Dr. Forbush is the copyright holder and developer of the Eating Pathology Symptoms Inventory.

Sara R. Gould, Danielle A. N. Chapa, Brittany K. Bohrer, Kelsey E. Hagan, Kelsey E. Clark, Daria A. Sorokina, and Victoria L. Perko each declare no potential conflicts of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:

•• Of major importance

- Stice E, Marti CN, Rohde P. Prevalence, incidence, impairment, and course of the proposed DSM-5 eating disorder diagnoses in an 8-year prospective community study of young women. *J Abnorm Psychol.* 2013;122:445–57.
- Allen KL, et al. DSM-IV-TR and DSM-5 eating disorders in adolescents: prevalence, stability, and psychosocial correlates in a population-based sample of male and female adolescents. *J Abnorm Psychol.* 2013;122:720–32.
- Arcelus J, et al. Mortality rates in patients with anorexia nervosa and other eating disorders: a meta-analysis of 36 studies. *Arch Gen Psychiatry.* 2011;68:724–31.
- Mitchell JE, Crow S. Medical complications of anorexia nervosa and bulimia nervosa. *Current Opinion in Psychiatry.* 2006;19:438–43.
- O'Brien KM, Vincent NK. Psychiatric comorbidity in anorexia and bulimia nervosa: nature, prevalence, and causal relationships. *Clin Psychol Rev.* 2003;23:57–74.
- Swanson SA, et al. Prevalence and correlates of eating disorders in adolescents: results from the national comorbidity survey replication adolescent supplement. *Arch Gen Psychiatry.* 2011;68:714–23.
- Fichter MM, Quadflieg N. Mortality in eating disorders—results of a large prospective clinical longitudinal study. *Int J Eat Disord.* 2016;49:391–401.
- Ackard DM, et al. Poor outcome and death among youth, young adults, and midlife adults with eating disorders: an investigation of risk factors by age at assessment. *Int J Eat Disord.* 2014;47:825–35.
- Mangweth-Matzek B, et al. Prevalence of eating disorders in middle-aged women. *Int J Eat Disord.* 2014;47:320–4.
- Marques L, et al. Comparative prevalence, correlates of impairment, and service utilization for eating disorders across US ethnic groups: implications for reducing ethnic disparities in health care access for eating disorders. *Int J Eat Disord.* 2011;44:412–20.
- Perez M, Ohrt TK, Hoek HW. Prevalence and treatment of eating disorders among Hispanics/Latino Americans in the United States. *Current Opinion in Psychiatry.* 2016;29:378–82.
- Pike KM, Hoek HW, Dunne PE. Cultural trends and eating disorders. *Current Opinion in Psychiatry.* 2014;27:436–42.
- Duncan AE, Ziobrowski HN, Nicol G. The prevalence of past 12-month and lifetime DSM-IV eating disorders by BMI category in US men and women. *Eur Eat Disord Rev.* 2017;25:165–71.
- Hudson JI, et al. The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biol Psychiatry.* 2007;61:348–58.
- Mulders-Jones B, et al. Socioeconomic correlates of eating disorder symptoms in an Australian population-based sample. *PLoS One.* 2017;12:e0170603.
- Turner H, Bryant-Waugh R, Marshall E. The impact of early symptom change and therapeutic alliance on treatment outcome in cognitive-behavioural therapy for eating disorders. *Behav Res Ther.* 2015;73:165–9.
- Hilbert A, et al. Rapid response in psychological treatments for binge eating disorder. *J Consult Clin Psychol.* 2015;83:649–54.
- Grilo CM, et al. Predicting meaningful outcomes to medication and self-help treatments for binge-eating disorder in primary care: the significance of early rapid response. *J Consult Clin Psychol.* 2015;83:387–94.
- Lambert MJ, Shimokawa K. Collecting client feedback. *Psychotherapy.* 2011;48(1):72.
- Hatfield D, et al. Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology & Psychotherapy.* 2010;17(1):25–32.
- Reese RJ, et al. Effect of client feedback on couple psychotherapy outcomes. *Psychother Theory Res Pract Train.* 2010;47(4):616.
- Lambert MJ, et al. Enhancing psychotherapy outcomes via providing feedback on client progress: a replication. *Clinical Psychology & Psychotherapy.* 2002;9(2):91–103.
- Finn SE. In our clients' shoes: theory and techniques of therapeutic assessment. 2007: Psychology Press.
- Yanovski SZ, et al. The Questionnaire on Eating and Weight Patterns-5 (QEWP-5): an updated screening instrument for binge eating disorder. *The International Journal of Eating Disorders.* 2015;48:259–61.
- Sandoz EK, et al. Assessment of body image flexibility: the body image-acceptance and action questionnaire. *Journal of Contextual Behavioral Science.* 2013;2:39–48.
- Martinsen M, et al. The development of the brief eating disorder in athletes questionnaire. *Med Sci Sports Exerc.* 2014;46:1666–75.
- Moher D, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7):e1000097.
- DeVellis RF. Scale development: theory and applications, vol. vol. 26. Newbury Park: Sage; 2016.
- Garner DM. EDI 3: eating disorder inventory-3: professional manual. Lutz: Psychological Assessment Resources; 2004.
- Moessner M, Fassnacht DB, Bauer S. Online assessment of eating disorders: the clinical and research inventory for eating disorders (CR-EAT). *Mental Health & Prevention.* 2015;3:170–7. **The CR-EAT is an online self-report measure with good reliability but problematic convergent and discriminant validity**
- Garner DM. Eating disorder inventory-2. Odessa: Psychological Assessment Resources; 1991.
- Bauer S, et al. Construction, scoring and validation of the Short Evaluation of Eating Disorders (SEED). *Eur Eat Disord Rev.* 2005;13(1):191–200.
- Killen JD, et al. Pursuit of thinness and onset of eating disorder symptoms in a community sample of adolescent girls: a three-year prospective analysis. *Int J Eat Disord.* 1994;16:227–38.
- Aardoom JJ, Dingemans AE, Van Furth EF. E-health interventions for eating disorders: emerging findings, issues, and opportunities. *Current Psychiatry Reports.* 2016;18:1–8.
- Tu C-Y, et al. Comparative validity of the Internet and paper-and-pencil versions of the Night Eating Questionnaire. *Compr Psychiatry.* 2017;75:53–61.
- Dias JCR, Maroco J, Campos JADB. Weight concerns scale applied to college students: comparison between pencil-and-paper and online formats. *Cyberpsychology, Behavior and Social Networking.* 2015;18(3):188–92.
- Alfonsson S, Maathz P, Hursti T. Interformat reliability of digital psychiatric self-report questionnaires: a systematic review. *J Med Internet Res.* 2014;16:e268.
- Tatham M, et al. Development, psychometric properties and preliminary clinical validation of a brief, session-by-session measure of eating disorder cognitions and behaviors: the ED-15. *Int J Eat Disord.* 2015;48:1005–15. **The ED-15 is a week-by-week self-**

- report measure with strong internal consistency and test-retest reliability but poor discriminant validity**
39. Fairburn CG. Cognitive behavior therapy and eating disorders. New York: Guilford; 2008.
 40. Spangler DL. The change in eating disorder symptoms scale: scale development and psychometric properties. *Eat Behav.* 2010;11(3): 131–7.
 41. Clark LA, Watson D. Constructing validity: basic issues in objective scale development. *Psychol Assess.* 1995;7:309–19.
 42. Sysko R, et al. The eating disorder assessment for DSM-5 (EDA-5): development and validation of a structured interview for feeding and eating disorders. *Int J Eat Disord.* 2015;48:452–63. **The EDA-5 is a semi-structured interview available as an app or in pencil-and-paper format with strong test-retest reliability, good concurrent validity for AN and BN and poor concurrent validity when no eating disorder diagnosis is appropriate. The EDA-5 also demonstrated adequate convergent and discriminant validity**
 43. APA. Diagnostic and statistical manual of mental disorders (DSM-5®). Arlington: American Psychiatric Publishers; 2013.
 44. Cooper Z, Fairburn C. The eating disorder examination: a semi-structured interview for the assessment of the specific psychopathology of eating disorders. *Int J Eat Disord.* 1987;6:1–8.
 45. Stice E, Telch CF, Rizvi SL. Development and validation of the Eating Disorder Diagnostic Scale: a brief self-report measure of anorexia, bulimia, and binge-eating disorder. *Psychol Assess.* 2000;12:123–31.
 46. Forbush KT, et al. Development and validation of the Eating Pathology Symptoms Inventory (EPSI). *Psychol Assess.* 2013;25: 859–78. **The EPSI is available online through Recovery Record, Inc. and in paper-and-pencil format. This measure has replicated factor structure and evidence for strong reliability and validity. Psychometric properties are sub-optimal for the Muscle Building scale in primarily female samples, but excellent in men**
 47. ter Huurne ED, et al. Is the Eating Disorder Questionnaire-Online (EDQ-O) a valid diagnostic instrument for the DSM-IV-TR classification of eating disorders? *Compr Psychiatry.* 2015;57:167–76. **The EDQ-O is an online self-report measure that corresponds to the DSM-IV eating disorder criteria. Reliability, convergent validity, and discriminant validity were not investigated. Concurrent validity was acceptable**
 48. APA. DSM-IV-TR: diagnostic and statistical manual of mental disorders, text revision. Washington: American Psychiatric Association; 2000. p. 75.
 49. Fairburn CG, Bèglin SJ. Assessment of eating disorders: interview or self-report questionnaire? *Int J Eat Disord.* 1994;16(4):363–70.
 50. Spitzer RL. Psychiatric diagnosis: are clinicians still necessary? *Compr Psychiatry.* 1983;24:399–411.
 51. Van Vliet I, Leroy H, Van Megen H. MINI Plus: MINI Internationaal Neuropsychiatrisch Interview: Dutch version 5.0. 0. LUMC: Leiden; 2000.
 52. Forbush KT, Wildes JE, Hunt TK. Gender norms, psychometric properties, and validity for the Eating Pathology Symptoms Inventory. *Int J Eat Disord.* 2014;47:85–91.
 53. Forbush KT, et al. Test–retest reliability of common measures of eating disorder symptoms in men versus women. *Assessment.* 2017; doi:10.1177/1073191117700267.
 54. Anderson DA, et al. Development and validation of a multifactorial treatment outcome measure for eating disorders. *Assessment.* 1999;6:7–20.
 55. Bohrer BK, Forbush KT, Hunt TK. Are common measures of dietary restraint and disinhibited eating reliable and valid in obese persons? *Appetite.* 2015;87:344–51.
 56. Tang X, Forbush KT, Lui PP. Development and validation of the Chinese-language version of the eating pathology symptoms inventory. *Int J Eat Disord.* 2015;48:1016–23.
 57. De Young KP, Anderson DA. An interactive, graphical tool for retrospectively assessing symptom frequency and severity an illustration with eating disorder behaviors, body weight, and stress. *Assessment.* 2016: 1073191116668629. **The IGAT is an online retrospective self-report measure with strong test-retest reliability and convergent validity overall, but both were lower for some constructs.**
 58. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol.* 1988;54(6):1063.
 59. Murray SB, Griffiths S, Mond JM. Evolving eating disorder psychopathology: conceptualising muscularity-oriented disordered eating. *Br J Psychiatry.* 2016;208:414–5.
 60. Wildes JE, Forbush KT. Ethnicity as a risk factor for eating disorders. In: *The Wiley handbook of eating disorders.* Wiley; 2015. p. 324–337.
 61. CDC. Vital signs: state-specific obesity prevalence among adults—United States, 2009. *MMWR Morb Mortal Wkly Rep.* 2010;59: 951–5.
 62. Loevinger J. Objective tests as instruments of psychological theory. *Psychol Rep.* 1957;3:635–94.
 63. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull.* 1959;56:81–105.